

## **MESTRADO**

# **MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E EMPRESARIAL**

## **TRABALHO FINAL DE MESTRADO**

### **RELATÓRIO DE ESTÁGIO**

**IDENTIFICAÇÃO E DELIMITAÇÃO DE TERRITÓRIOS  
HOMOGÉNEOS DE RESIDENTES COM NÍVEIS DE  
RENDIMENTO E PATRIMÓNIO MAIS ELEVADOS**

**CATARINA ISABEL FERNANDES VALENTE**

**DEZEMBRO - 2019**

**MESTRADO EM**  
**MÉTODOS QUANTITATIVOS PARA A DECISÃO**  
**ECONÓMICA E EMPRESARIAL**

**TRABALHO FINAL DE MESTRADO**  
**RELATÓRIO DE ESTÁGIO**

IDENTIFICAÇÃO E DELIMITAÇÃO DE TERRITÓRIOS  
HOMOGÉNEOS DE RESIDENTES COM NÍVEIS DE  
RENDIMENTO E PATRIMÓNIO MAIS ELEVADOS

CATARINA ISABEL FERNANDES VALENTE

**ORIENTAÇÃO:**

DR. FRANCISCO VALA  
PROFESSOR DR. PAULO PARENTE

DEZEMBRO - 2019

### Agradecimentos

Em primeiro lugar, quero agradecer de um modo particular aos meus orientadores Dr. Francisco Vala e Prof. Paulo Parente por todo o acompanhamento e tempo dedicado a este trabalho.

Agradeço à minha mãe o apoio, investimento, dedicação, compreensão e educação ao longo do meu percurso académico.

Ao INE pela oportunidade de estágio, em especial à equipa do Gabinete para a Coordenação de Estatísticas Territoriais pela calorosa receção e integração, pelo acompanhamento e partilha de conhecimentos ao longo do estágio, destacando a colaboração e disponibilidade do colega Rossano Figueiredo.

À minha colega de estágio e mestrado Patrícia Rocha pela disponibilidade e ajuda neste trabalho.

Por fim, à minha irmã, ao meu namorado e amigos pela amizade, incentivo e motivação demonstrados.

### Resumo

Em inquéritos associados à temática da riqueza, como é o caso do Inquérito à Situação Financeira das Famílias (ISFF), é particularmente relevante assegurar um desenho amostral que garanta a representação da população para todos os escalões de riqueza. Contudo, a elevada assimetria da distribuição da riqueza contribui como um obstáculo para este objetivo, uma vez que tende a ser menor a probabilidade de ser selecionada, para a amostra, a população onde se encontra a maior parte da riqueza. Uma forma de ultrapassar este entrave, traduz-se numa sobreamostragem deste grupo. Neste contexto, o presente relatório de estágio tem como objetivo identificar e delimitar territórios homogéneos de residência de população com níveis de rendimento e património potencialmente mais elevados. Assim, foram criados diversos modelos com base em análises de cluster e fatorial, recorrendo a variáveis de contexto socioeconómico, de rendimento e associadas ao mercado imobiliário, à escala local.

Palavras-chave: HFCS, ISFF, rendimento, amostra, cluster, análise fatorial

### Abstract

In wealth surveys, such as the portuguese Household Financial and Consumption Survey (called Inquérito à Situação Financeira das Famílias - ISFF), it is relevant to have a sample design that ensures population representation for all wealth levels. However, a high asymmetric distribution of wealth contributes as a barrier to this goal since the population which includes a large portion of wealth is less likely to be selected for the sample. One way to overcome this difficulty is oversampling this group. In this context, this internship report aims to identify and delimit homogeneous regions of residence of population with potentially higher income and assets levels. Thus, several models were created based on cluster and factorial analysis, using variables of socioeconomic context, income and real estate market, on a local scale.

Keywords: HFCS, ISFF, income, sample, cluster, factor analysis

## Índice

<b>Agradecimentos.....</b>	<b>3</b>
<b>Resumo.....</b>	<b>4</b>
<b>Abstract.....</b>	<b>5</b>
<b>Lista de Figuras.....</b>	<b>7</b>
<b>Lista de Tabelas .....</b>	<b>10</b>
<b>Acrónimos .....</b>	<b>11</b>
<b>1.Introdução .....</b>	<b>12</b>
<b>2.Revisão da Literatura.....</b>	<b>14</b>
<b>3. Dados e Metodologia .....</b>	<b>19</b>
3.1. Rendimento .....	19
3.2. Preços da habitação.....	20
3.3. Rendas da habitação.....	21
3.4. Censos 2011 .....	23
3.5. Agregação de freguesias .....	24
3.6. Análise Univariada .....	26
3.7. Análise Bivariada.....	27
3.8. Análise fatorial.....	29
3.9. Análise de clusters .....	30
<b>4.Resultados .....</b>	<b>34</b>
4.1. Estudo 1 .....	35
4.2. Estudo 2 .....	36
4.3. Estudo 3 .....	37
4.4. Estudo 4 .....	39
4.5. Análise de Resultados .....	40
<b>5.Conclusão .....</b>	<b>41</b>
<b>Referências Bibliográficas.....</b>	<b>42</b>
<b>Anexos .....</b>	<b>44</b>

### Lista de Figuras

Figura 1: Divisão administrativa por freguesia, 2014 .....	26
Figura 2: Agregação de freguesias segundo a TIPAU e os preços de habitação, 2014 e 2017.....	26
Figura 3: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 1.....	35
Figura 4: Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 1, por região, 2016.....	35
Figura 5: Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 1, por região, 2016.....	35
Figura 6: Percentil 80 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por região, 2017.....	36
Figura 7: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por região, 2017.....	36
Figura 8: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 2.....	36
Figura 9: Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 2, por região, 2016.....	37
Figura 10: Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 2, por região, 2016.....	37
Figura 11: Percentil 90 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por região, 2017 .....	37
Figura 12: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por região, 2017.....	37
Figura 13: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 3.....	38
Figura 14: Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 3, por região, 2016.....	38

Figura 15: Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 3, por região, 2016.....	38
Figura 16: Percentil 80 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o 3, por região, 2017.....	38
Figura 17: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 3, por região, 2017.....	38
Figura 18: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 4.....	39
Figura 19: Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 4, por região, 2016.....	39
Figura 20: Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 4, por região, 2016.....	39
Figura 21: Percentil 90 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 4, por região, 2017.....	40
Figura 22: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 4, por região, 2017.....	40
Figura 23: Agregação de freguesias.....	45
Figura 24: Matriz de correlação das variáveis por região.....	46
Figura 25: Matriz de correlação das variáveis por freguesia.....	46
Figura 26: Classificação das freguesias para três clusters segundo o Estudo 1...	47
Figura 27: Classificação das freguesias para três clusters segundo o Estudo 2...	47
Figura 28: Classificação das freguesias para três clusters segundo o Estudo 3...	47
Figura 29: Classificação das freguesias para três clusters segundo o Estudo 4...	47
Figura 30: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 1.....	51
Figura 31: Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 1, por freguesia, 2016.....	51



Figura 32: Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 1, por freguesia, 2016.....	51
Figura 33: Percentil 80 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por região, 2017.....	51
Figura 34: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por freguesia, 2017.....	51
Figura 35: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 2.....	52
Figura 36: Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 2, por freguesia, 2016.....	52
Figura 37: Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 2, por freguesia, 2016.....	52
Figura 38: Percentil 90 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por freguesia, 2017.....	52
Figura 39: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por freguesia, 2017.....	52
Figura 40: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 3.....	53
Figura 41: Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 3, por freguesia, 2016.....	53
Figura 42: Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 3, por freguesia, 2016.....	53
Figura 43: Percentil 80 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o 3, por freguesia, 2017.....	53

Figura 44: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 3, por freguesia, 2017.....	53
Figura 45: Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 4.....	54
Figura 46: Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 4, por freguesia, 2016.....	54
Figura 47: Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 4, por freguesia, 2016.....	54
Figura 48: Percentil 90 dos preços por m <sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 4, por freguesia, 2017.....	54
Figura 49: Proporção de alojamentos familiares com o preço por m <sup>2</sup> acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 4, por freguesia, 2017.....	54
Figura 50: Classificação das freguesias para três clusters segundo o Estudo 1...	55
Figura 51: Classificação das freguesias para três clusters segundo o Estudo 2...	55
Figura 52: Classificação das freguesias para três clusters segundo o Estudo 3...	55
Figura 53: Classificação das freguesias para três clusters segundo o Estudo 4...	55

#### Lista de Tabelas

Tabela I: Distribuição da segunda subamostra de acordo com a zona geográfica e respetiva classe de área útil.....	44
Tabela II: Códigos e designações da divisão administrativa para o 2º e 3º níveis (municípios e freguesias, respetivamente) de acordo com o cluster com um nível mais elevado de riqueza do estudo 3.....	48

### Acrónimos

AP – Area-Probability

AT – Autoridade Tributária e Aduaneira

BCE – Banco Central Europeu

BCN – Bancos Centrais Nacionais da Europa

SCF – Survey Consumer Finances

ECB – European Central Bank

FNA – Fichero Nacional de Alojamentos

FRB – Federal Reserve Bulletin

HFCS – Household Finance and Consumption Survey

HFCN – Household Finance and Consumption Network

INE – Instituto Nacional de Estatística

IRS – Imposto sobre o Rendimento das Pessoas Singulares

ISFF – Inquérito à Situação Financeira da Famílias

NUTS II – Nomenclatura das Unidades Territoriais para Fins Estatísticos, nível dois

PSU – Primary Sampling Units

SSU – Secondary Sampling Units

SSE - Sum of Squared Errors

## 1.Introdução

A distribuição da riqueza caracteriza-se pela sua elevada assimetria, ou seja, uma grande percentagem da riqueza encontra-se numa percentagem muito reduzida da população. Este facto contribui para que a probabilidade deste grupo pertencer à amostra de inquéritos associados à riqueza seja menor. Além disso, as famílias mais ricas também apresentam uma menor disponibilidade para responder a este tipo de inquéritos. Neste contexto, surge a necessidade de encontrar um desenho amostral representativo da população para todos os escalões de riqueza.

No âmbito de um projeto europeu designado por Household Finance and Consumption Survey (HFCS), foi criado o Inquérito à Situação Financeira das Famílias (ISFF) com o objetivo de recolher informação sobre a situação financeira das famílias. Neste inquérito, procurando uma representação adequada da distribuição da riqueza na população, foram explorados diversos métodos de forma a que no desenho da amostra exista uma sobrerrepresentação das famílias com um maior nível de riqueza. Esta problemática motivou a realização deste estudo.

O presente relatório tem como objetivo identificar e delimitar territórios homogêneos de residência de população com níveis de rendimento e património potencialmente mais elevados. A abordagem deste tema passou pela construção de uma base de dados de variáveis ao nível da freguesia, aplicando técnicas uni e multivariada (análise fatorial e de clusters). Para tal, foram usadas variáveis de contexto socioeconómico, de rendimento e associadas ao mercado imobiliário. Salienta-se que, tendo em conta o reduzido número de dados em algumas freguesias e de forma a evitar a sua exclusão desta análise, foi necessário criar uma nova unidade geográfica através da agregação de freguesias.

Este relatório de estágio enquadra-se no contexto do Mestrado Métodos Quantitativos para a Decisão Económica e Empresarial do Instituto Superior de Economia e Gestão, da Universidade de Lisboa. O enfoque deste documento é descrever as atividades desenvolvidas ao longo do estágio, assim como apresentar o enquadramento do trabalho realizado. O estágio realizou-se no Instituto Nacional de Estatística (INE), no Gabinete para a Coordenação das Estatísticas Territoriais, no período de 11 de Março a 2 de Agosto de 2019.

O INE como uma autoridade estatística independente, imparcial e credível, que desenvolve processos estatísticos metodologicamente avançados, tem por missão produzir e divulgar informação estatística oficial de qualidade, relevante para a sociedade, promovendo a coordenação, a análise, a inovação e a divulgação da atividade estatística nacional, garantindo o armazenamento integrado de dados e assegurando o respeito pela confidencialidade dos cidadãos e entidades.

A estrutura deste documento encontra-se dividida em cinco capítulos. No segundo capítulo é apresentada a revisão da literatura, onde são expostas as metodologias aplicadas na seleção da segunda subamostra, em edições anteriores do HFCS (dos diversos países participantes) e a metodologia aplicada num inquérito semelhante (Survey Consumer Finances) realizado no Estados Unidos da América. No terceiro capítulo, encontram-se descritas as fontes e a metodologia utilizadas para a construção da base de dados, assim como para a análise de clusters e fatorial. No quarto capítulo, apresenta-se as observações dos resultados obtidos, seguindo-se, no quinto capítulo, as conclusões do estudo, referindo as suas limitações e sugestões de futuras investigações.

## 2.Revisão da Literatura

Em 2008, o Conselho do Banco Central Europeu (BCE) criou um projeto denominado por Household Finance and Consumption Survey (HFCS), que se traduz num inquérito que visa recolher informação que permita caracterizar detalhadamente a situação financeira das famílias. As principais questões incidem sobre o património das famílias, nomeadamente sobre os ativos não financeiros que estas possuam, sobre eventuais empréstimos que utilizem esses ativos como garantia, outras responsabilidades financeiras e sobre as aplicações financeiras da família. A informação recolhida ao nível da família completa-se com questões sobre heranças eventualmente recebidas, rendimentos (provenientes de negócios, de investimentos imobiliários e financeiros e ainda de transferências) e sobre as suas decisões de consumo e poupança. Além disso, o HFCS também contribui para a realização de estudos no contexto da política monetária e da manutenção da estabilidade financeira neste grupo de países (Costa, S. e L. Farinha, 2012).

Neste contexto, foi criado um grupo de trabalho designado Household Finance and Consumption Network (HFCN). Este grupo é composto por Bancos Centrais Nacionais da Europa (BCN), alguns Institutos Nacionais de Estatística e vários outros especialistas, tendo como propósito desenvolver e realizar o HFCS. A realização do HFCS é descentralizada a nível nacional, sendo da responsabilidade de cada instituição participante. O Banco Central Europeu (BCE), em conjunto com o HFCN, coordena todo o projeto, garantindo a aplicação de uma metodologia comum, agrupando e controlando a qualidade dos conjuntos de dados de cada país.

Em inquéritos às famílias, como é o caso do HFCS, a técnica de seleção de amostra mais utilizada é a de estratificação da população. Esta técnica caracteriza-se por permitir obter estimativas mais precisas da população, no caso da variável em estudo assumir valores médios diferentes, em diferentes subpopulações (Lohr, 2010, p.74). Além disso, numa amostra estratificada, a seleção da amostra pode ser realizada em diferentes etapas, sendo que, neste contexto, a primeira corresponde à seleção da área geográfica e a última à escolha do alojamento onde a família será inquirida.

Na primeira edição do HFCS, sete países utilizaram apenas uma etapa e os restantes duas ou três. (ECB, 2013) Portugal estratificou a sua amostra em duas etapas. Na primeira

foram selecionadas unidades primárias (PSU – Primary Sampling Units<sup>1</sup>) com probabilidade proporcional à dimensão dos alojamentos de residência principal, para cada uma das regiões, e na segunda etapa, unidades secundárias (SSU – Secondary Sampling Units, os alojamentos) selecionadas de forma sistemática dentro das unidades da primeira etapa. A partir da Amostra-Mãe<sup>2</sup> de 2001 foram criadas duas subamostras: a primeira corresponde a uma “amostra aleatória representativa dos agregados familiares residentes em alojamentos não coletivos ao nível nacional, estratificada por NUTS II<sup>3</sup>” (Banco de Portugal e INE, 2012), baseando-se na lista de áreas geográficas e de habitações privadas registadas nos Censos e a segunda à sobreamostragem dos agregados familiares com rendimento e património potencialmente mais elevados.

A distribuição assimétrica da riqueza não é exclusiva de Portugal, por isso todos os países participantes no HFCS têm sido incentivados a realizar uma sobreamostragem dos agregados familiares mais ricos. Assim, quanto maior for o rigor na recolha dos dados, maior é a garantia de que os valores estimados da riqueza total não se encontram enviesados. Dos quinze países intervenientes, apenas nove realizaram a sobreamostragem. Diversas foram as estratégias aplicadas na recolha de dados que pudessem identificar as famílias de interesse a inquirir. Também a disponibilidade dos dados em cada país contribuiu para estas diferenças. Na primeira edição, Espanha recorreu aos dados individuais referentes à riqueza tributável, França à riqueza líquida individual enquanto que a Finlândia e o Luxemburgo utilizaram os valores do nível de rendimento por indivíduo. Noutros países, utilizaram-se diferentes dados relacionadas com riqueza. Chipre utilizou o consumo de eletricidade por agregado familiar, a Bélgica e a Alemanha o nível de rendimento por região e a Grécia os preços imobiliários. Por fim, os restantes países tomaram como base as regiões ou municípios onde tinham sido observados valores de rendimento mais elevados e/ou eram esperadas taxas de não-resposta superiores. Portugal, pertencendo a este grupo, reforçou a sua amostra de base

---

<sup>1</sup> “Áreas constituídas por uma ou mais células quadradas com 1km de lado da grelha europeia INSPIRE” (INE, 2016).

<sup>2</sup> “A Amostra-Mãe (AM) é uma amostra probabilística, representativa dos alojamentos familiares de residência principal ao nível das NUTS II, sendo constituída por 1408 áreas geográficas. Esta amostra foi criada a partir dos Censos de 2001, sendo atualizada permanentemente. Será substituída quando forem utilizados para este efeito os resultados dos Censos de 2011” (Costa & Farinha, 2012).

<sup>3</sup> NUTS II (Nomenclatura das Unidades Territoriais para Fins Estatísticos, nível dois) corresponde a sete divisões regionais: Norte, Centro; Área Metropolitana de Lisboa, Alentejo; Algarve, Açores e Madeira.

das áreas Metropolitanas de Lisboa e Porto em 50%, para a sobreamostragem. (ECB,2013)

O Banco Central Europeu apresenta no seu relatório metodológico relativo à primeira edição do HFCS (2013) um indicador da representação da riqueza na amostra final: a taxa efetiva de sobreamostragem da riqueza. Esta taxa indica o quanto a quota de agregados familiares mais ricos na amostra excede a sua quota na população. Este indicador é apresentado de acordo com os dez e cinco por cento dos agregados mais ricos na população. Primeiramente, calcula-se o percentil 90 ou 95 dos dados ponderados da riqueza líquida que será utilizado como referência para o cálculo da percentagem de agregados familiares mais ricos na amostra e na população. Sendo S90 a quota de agregados com maior riqueza acima do percentil 90, tem-se  $\frac{S90-0.1}{0.1}$  como a taxa efetiva de sobreamostragem da riqueza para os 10% de agregados familiares com maior riqueza. No caso de S95 (quota de agregados com maior riqueza acima do percentil 95), tem-se  $\frac{S95-0.05}{0.05}$  como a taxa efetiva de sobreamostragem da riqueza para os 5% de agregados familiares com maior riqueza. Assim, se a quota(S90) for igual a 10%, a taxa para o top 10% é igual a zero; se S90 for acima de 10%, a taxa é positiva e se S90 for abaixo de 10%, a taxa é negativa. A mesma interpretação é aplicada para o top 5%.

Para o top 10%, destacam-se com maiores valores Espanha (192%), França (129%) e Alemanha (117%). Para o top 5%, salientam-se os mesmos países: Espanha (314%), França (208%) e Alemanha (148%). Em qualquer dos dois casos, Portugal apresenta valores positivos (16% e 20%, respetivamente), mas ainda muito distantes destes países. (ECB,2013)

De acordo com o BCE (2016), na segunda edição, também diferentes critérios de estratificação foram aplicados. Cinco países utilizaram uma etapa de estratificação enquanto os outros aplicaram múltiplas ou nenhuma. Portugal recorreu a um esquema de amostragem estratificado e multietápico com duas etapas de estratificação (PSU e SSU). Ambas as amostras foram extraídas do Ficheiro Nacional de Alojamentos (FNA) que já engloba todos os alojamentos do território nacional e contém informações mais atualizadas, visto terem como fonte os Censos 2011. A primeira subamostra foi criada com base em nove regiões correspondentes a NUTS II com desagregação do Norte em “Porto” e “Norte exceto Porto” e da Área Metropolitana de Lisboa em “Grande Lisboa”



e “Península de Setúbal”. Tendo sido concluído, com base nos dados da primeira edição do ISFF, que a dimensão dos alojamentos está mais correlacionada com o nível de riqueza do que com a localização geográfica, a esta seleção foram associadas duas classes de área útil do alojamento (uma classe inferior ao valor e outra superior ou igual ao valor), para cada um dos três valores de referência<sup>4</sup> (100m<sup>2</sup>, 120m<sup>2</sup>, 150m<sup>2</sup>, consoante as regiões). A segunda subamostra diz respeito à sobrerepresentação dos agregados familiares que apresentem valores de riqueza mais elevados, detalhada mais à frente.

Nesta edição do HFCS (2013) em que participaram vinte países, quinze realizaram uma sobreamostragem dos alojamentos com maior riqueza. Destes quinze países fazem parte aqueles que na primeira edição já aplicaram uma sobreamostragem, a Eslováquia e os novos participantes desta edição. Relativamente às estratégias de sobreamostragem aplicadas, Espanha recolheu informação dos dados individuais da riqueza tributável, França dos dados da riqueza líquida individual e Estónia, Finlândia, Letónia e Luxemburgo do rendimento individual. Chipre utilizou como variável aproximada o consumo de eletricidade ao nível das famílias. Os restantes países, devido à falta de acesso a informação mais adequada, utilizaram dados relacionados com o rendimento a nível regional, preços de habitação e/ou a condição laboral a nível individual. (ECB,2016) Em Portugal, a sobreamostragem consistiu em alojamentos pertencentes, em cada uma das nove regiões anteriormente referidas, à classe de área útil de maior dimensão ( $\geq 100\text{m}^2$  na região “Grande Lisboa”,  $\geq 120\text{m}^2$  nas regiões “Porto”, “Península de Setúbal”, “Algarve” e “Região Autónoma da Madeira” e  $\geq 150\text{m}^2$  em “Norte exceto Porto”, “Centro”, “Alentejo” e “Região Autónoma dos Açores” – Anexo A, tabela I). De acordo com a taxa efetiva de sobreamostragem da riqueza, Espanha (234%, 374%), Alemanha (141%, 173%) e França (132%, 227%) foram os países que apresentaram os resultados mais altos, no top 10% e 5%, respetivamente. Comparativamente à primeira edição, a maioria dos países obteve resultados semelhantes ou ligeiramente superiores, destacando-se a Eslováquia (que aumentou dezasseis pontos percentuais, atingindo 5% para o top 10% e vinte e três pontos percentuais, atingindo 15% para o top 5%) e Portugal (que

---

<sup>4</sup> Limites definidos com base na informação recolhida no ISFF 2010, de acordo com as diferenças regionais em termos de dimensão de alojamento.

aumentou trinta e sete pontos percentuais, atingindo 53% para o top 10% e trinta e um pontos percentuais, atingindo 51% para o top 5%). (ECB,2016)

Com base nestes resultados, pôde-se verificar que as estratégias aplicadas à sobreamostragem, assim como os dados disponíveis a nível do alojamento desempenham um papel fundamental no que diz respeito à eficácia da obtenção de respostas dos alojamentos mais ricos. (ECB, 2016)

Tendo em conta que os resultados da terceira edição ainda não foram publicados, pode-se apenas perceber, através do documento metodológico versão 1.2 (INE, 2017), que a metodologia aplicada em Portugal relativamente ao desenho amostral não foi alterada.

Em muitos outros países, além dos países da área do euro existem inquéritos semelhantes ao HFCS. Um exemplo pioneiro destes inquéritos consiste no SCF (Survey Consumer Finances), conduzido pela Reserva Federal Norte-Americana desde 1983. Em 1989, foi realizada uma profunda reavaliação deste inquérito e do seu desenho amostral, que desde então têm sido sujeitos a apenas ligeiras alterações, mantendo-os quase idênticos. O SCF aplica duas técnicas de amostragem aleatória. A primeira consiste numa amostra padrão de probabilidade de base geográfica (AP – area-probability) e a segunda à sobre-representação dos agregados familiares com maior riqueza (“list sample”). A amostra AP é selecionada em três etapas. Na primeira é aplicado um esquema de estratificação para a escolha das áreas metropolitanas e os municípios rurais. Na segunda, o mesmo esquema é utilizado para a seleção das subáreas. Por fim, na última, os alojamentos são escolhidos através de uma amostra sistemática, com base na lista de endereços das subáreas, de forma a garantir igual probabilidade de seleção na amostra geral. (FRB,2014; Kennickell, 2007)

De acordo com Federal Reserve Bulletin (FRB, 2014), a “list sample” consiste num grupo desenhado a partir de uma lista de dados oriundos das declarações de rendimentos e impostos, devidamente editada. Esta seleção é realizada em duas etapas. Na primeira são selecionadas observações da AP e na segunda é aplicada uma estratificação em que uma parte modela a riqueza como uma simples extrapolação dos fluxos de rendimento do capital e a outra é estimada com base numa correspondência anónima entre a amostra e os dados das edições anteriores do SCF.

### 3. Dados e Metodologia

De modo a identificar e classificar os territórios homogêneos de residência da população com rendimentos e património mais elevados, foi construída uma base de dados recorrendo aos microdados de quatro fontes principais: a informação associada às notas de liquidação do IRS 2016, a base de dados subjacente às Estatísticas de Preços da Habitação ao nível local 2016 e 2017, a base de dados subjacente às Estatísticas de Rendas da Habitação ao nível local 2016 e 2017 e as variáveis censitárias 2011. Uma vez construída a base de dados, verificou-se que grande parte das freguesias não apresentavam dados para muitas das variáveis calculadas, impossibilitando uma análise que cobrisse todo o território nacional. Para ultrapassar este obstáculo, foi criada uma nova composição geográfica composta por freguesias e agregações de freguesias. De seguida, procedeu-se à análise uni e bivariada que contribuiu para a seleção das variáveis. Por fim, realizaram-se as análises fatorial e de clusters para identificação e classificação dos territórios. Para este estudo recorreu-se à utilização de *softwares* como Rstudio e SPSS.

#### 3.1. Rendimento

A informação associada às notas de liquidação do Imposto sobre o rendimento das pessoas singulares 2016 (IRS – Modelo 3) resulta de fontes administrativas cujos dados fiscais (que abrangem o país) são anonimizados pela Autoridade Tributária e Aduaneira (AT). É de salientar que “na nota de liquidação do IRS não é possível distinguir nem a origem e natureza dos rendimentos abrangidos” (INE, 2019b) nem a composição dos agregados familiares. Para o cálculo das variáveis, considerou-se apenas os agregados fiscais com rendimento bruto (ou global) declarado deduzido do IRS liquidado maior que zero. Assim, teve-se por base o valor anual do rendimento bruto declarado (composto pelos rendimentos declarados e que constituem a base de incidência do imposto: rendimentos do trabalho dependente, empresariais e profissionais, de capitais, prediais, incrementos patrimoniais e pensões) e a coleta líquida (imposto a pagar de IRS, após consideradas todas as deduções) para o cálculo do valor anual do rendimento líquido (rendimento bruto declarado deduzido de IRS liquidado). Deste modo, foram apuradas as seguintes variáveis referentes a cada freguesia:

- os percentis 50,75,80 e 90 do rendimento global/líquido;

- o coeficiente de variação do rendimento global/líquido;
- a proporção de agregados familiares com o rendimento global/líquido superior ao percentil 50/75/80/90 a nível nacional<sup>5</sup>.

### 3.2. Preços da habitação

As Estatísticas de Preços da Habitação, segundo o seu documento metodológico (INE, 2018), são um estudo estatístico que, com base em fontes administrativas, “recorre a informação complementar da Agência para a Energia (ADENE), nomeadamente, as coordenadas geográficas de localização dos alojamentos familiares, para além de informação administrativa fiscal do Imposto Municipal sobre Transmissões onerosas de imóveis (IMT) e do Imposto Municipal sobre Imóveis (IMI)” (INE, 2018). Esta operação estatística de periodicidade trimestral, abrange todo o país e tem como objetivo disponibilizar informação de interesse público e académico, sendo possível acompanhar a dinâmica do mercado de venda de alojamentos familiares destinados à habitação. As variáveis associadas aos ficheiros administrativos são sujeitas, *a priori*, a um conjunto de regras de validação quer por parte da AT quer do INE. Para o cálculo dos resultados, é criada uma ligação entre a informação (do IMT) referente aos preços das transações e a informação (do IMI) sobre as características do alojamento transacionado. Além disso, teve-se em consideração algumas restrições. Desta forma, foram excluídos “os contratos de troca ou permuta de bens imóveis”, “as declarações com valor igual a 0€”, “as transações com códigos de destino e afetações diferentes de habitação” e “os imóveis com área bruta privativa inferior a 20m<sup>2</sup> e superior ou igual a 600m<sup>2</sup>”. Assim, tendo em conta os preços de habitação em 2016 e 2017, foram calculadas as seguintes variáveis para cada freguesia:

- os percentis 50, 75, 80 e 90 do preço por m<sup>2</sup>;
- os percentis 50, 75, 80 e 90 do valor total da transação;
- os percentis 50, 75, 80 e 90 da área total do alojamento familiar;
- os percentis 50, 75, 80 e 90 do preço por m<sup>2</sup> dos alojamentos familiares do tipo apartamento/moradia;

---

<sup>5</sup> Número de agregados familiares com o rendimento global/líquido superior ao respetivo percentil a nível nacional, em relação ao número total de agregados familiares, por freguesia, multiplicado por 100.

- os percentis 50, 75, 80 e 90 do preço por m<sup>2</sup> dos alojamentos familiares novos/existentes;
- o coeficiente de variação do preço por m<sup>2</sup>;
- o coeficiente de variação da área do alojamento familiar;
- o coeficiente de variação do valor total transacionado;
- a proporção de alojamentos familiares com a tipologia igual ou superior a T4<sup>6</sup>;
- a proporção de alojamentos familiares com o preço por m<sup>2</sup> superior ao respetivo percentil 50/75/80/90, a nível nacional<sup>7</sup>;
- a proporção de alojamentos familiares com o valor total da transação superior ao respetivo percentil 50/75/80/90, a nível nacional<sup>8</sup>;
- a proporção de alojamentos familiares com a área total do alojamento superior ao respetivo percentil 50/75/80/90, a nível nacional<sup>9</sup>;
- a proporção de alojamentos familiares do tipo apartamento/moradia, com o preço por m<sup>2</sup> superior ao respetivo percentil 50/75/80/90, a nível nacional<sup>10</sup>;
- a proporção de alojamentos familiares novos/existentes, com o preço por m<sup>2</sup> superior ao respetivo percentil 50/75/80/90, a nível nacional<sup>11</sup>;

### 3.3. Rendas da habitação

Também as Estatísticas de Rendas da Habitação (descritas no respetivo documento metodológico – INE, 2019a) são um estudo estatístico cujos dados são obtidos através de fontes administrativas. Esta operação estatística abrange todo o território nacional e tem uma periodicidade semestral. Tendo em vista o interesse público e académico e com base em “informação administrativa fiscal do Imposto municipal sobre imóveis (IMI) e do

---

<sup>6</sup> Número de alojamentos vendidos com uma tipologia igual ou superior a T4 em relação ao total de alojamentos vendidos, multiplicado por 100.

<sup>7</sup> Número de alojamentos vendidos com o preço por m<sup>2</sup> superior ao respetivo percentil a nível nacional, em relação ao número total de alojamentos vendidos, por freguesia, multiplicado por 100.

<sup>8</sup> Número de alojamentos vendidos com o valor total da transação superior ao respetivo percentil a nível nacional, em relação ao número total de alojamentos vendidos, por freguesia, multiplicado por 100.

<sup>9</sup> Número de alojamentos vendidos com a área total do alojamento superior ao respetivo percentil a nível nacional, em relação ao número total de alojamentos vendidos, por freguesia, multiplicado por 100.

<sup>10</sup> Número de alojamentos vendidos do tipo apartamento/moradia, com o preço por m<sup>2</sup> superior ao respetivo percentil a nível nacional, em relação ao número total de alojamentos vendidos, por freguesia, multiplicado por 100.

<sup>11</sup> Número de alojamentos novos/existentes, com o preço por m<sup>2</sup> superior ao respetivo percentil a nível nacional, em relação ao número total de alojamentos vendidos, por freguesia, multiplicado por 100.

Imposto do selo – Comunicação de contratos de arrendamento (Modelo 2)”, este estudo permite acompanhar a dinâmica do mercado de arrendamento de alojamentos familiares destinados à habitação permanente. As variáveis associadas aos ficheiros administrativos referidos foram sujeitas a um conjunto de regras de validação por parte da AT, bem como a um tratamento de dados e a uma análise realizados pelo INE. Para o cálculo destas estatísticas criou-se ainda uma ligação entre a informação dos preços dos novos contratos de arrendamento (informação do Modelo 2 do Imposto de Selo) e as características referentes ao alojamento arrendado (informação do IMI). Foram ainda aplicadas restrições que levaram a incluir apenas a “1ª Declaração” e “as declarações de substituição” desde que alterem algum dos campos que incidem no meu estudo e a excluir:

- ◇ “as alterações e as cessações de contrato”,
- ◇ “os prédios rústicos”, “os registos com período de renda inferior a um mês”,
- ◇ “as finalidades “Não habitacional” e “Habitacional (não permanente)””,
- ◇ “os subarrendamentos e as promessas de arrendamentos com entrega do bem locado”,
- ◇ os artigos matriciais não afetos à “Habitação”,
- ◇ os alojamentos com mais de dois novos contratos de arrendamento nos últimos 12 meses associados ao período de referência,
- ◇ os contratos de arrendamento para habitação com área bruta privativa inferior a 20m<sup>2</sup> e superior ou igual a 600m<sup>2</sup> <sup>12</sup>;
- ◇ os contratos de arrendamento com valor de renda inferior a 100€ e superior a 4000€<sup>13</sup>.

Assim, os dados referentes aos novos contratos de arrendamento realizados em 2016 e 2017 permitiram o cálculo das seguintes variáveis, ao nível da freguesia:

- os percentis 50,75, 80 e 90 da renda por m<sup>2</sup>;

---

<sup>12</sup> Relativamente a esta restrição nas Estatísticas de preços da habitação e nas Estatísticas de rendas da habitação, verifica-se que “O limite superior de 600 m2 permite a exclusão de arrendamentos de alojamentos que pela sua dimensão constituem exceções no mercado de arrendamento, podendo influenciar o valor por metro quadrado da unidade territorial” (INE,2018; INE, 2019a).

<sup>13</sup> “Esta opção resulta da análise da distribuição do valor dos novos contratos de arrendamento registados, na fase de avaliação da viabilidade desta operação estatística.” (INE, 2019a).

- a proporção de alojamentos arrendados com um coeficiente de qualidade e conforto superior ao seu valor mediano nacional (1%)<sup>14</sup>;
- a proporção de alojamentos arrendados classificados como moradias unifamiliares<sup>15</sup>;
- a proporção de alojamentos arrendados com a renda por m<sup>2</sup> superior ao respetivo percentil 50/75/80/90, a nível nacional<sup>16</sup>.

### 3.4. Censos 2011

Os XV Recenseamento Geral da População e V Recenseamento Geral da Habitação, também designados por Censos 2011, fazem parte de uma série estatística de recenseamentos realizados desde 1970. São caracterizados por serem efetuados com uma periodicidade decenal e por terem como principal objetivo “a recolha, o apuramento, a análise e a divulgação de dados estatísticos oficiais referentes às características demográficas e socioeconómicas da população residente em Portugal e do parque habitacional” (INE, 2010). Além disso, têm em vista a criação de uma base de informação de referência (de onde são extraídas amostras para inquéritos) e a organização de uma base de dados. Dada a importância dos censos, todo o processo e tratamento de dados foi cuidadosamente estudado, de forma a produzir resultados de interesse nacional e comparáveis internacionalmente. A recolha para os Censos 2011 foi realizada de forma direta (entrega e recolha de questionários porta a porta) ou, em alguns casos, através da internet (e-censos – opção disponibilizada pela primeira vez nesta edição). Tendo em conta as quatro unidades estatísticas observadas (edifícios, alojamentos, famílias e indivíduos) considerou-se relevante a informação sobre os alojamentos e indivíduos que permitiram o cálculo das seguintes variáveis por freguesia, referente a 2011:

- a proporção de alojamentos familiares cujo residente era proprietário ou coproprietário da habitação<sup>17</sup>;

---

<sup>14</sup> Número alojamentos arrendados com um coeficiente de qualidade e conforto superior ao seu valor mediano nacional (1%), em relação ao número total de alojamentos arrendados, por freguesia, multiplicado por 100.

<sup>15</sup> Número alojamentos arrendados classificados como moradias unifamiliares, em relação ao número total de alojamentos arrendados, por freguesia, multiplicado por 100.

<sup>16</sup> Número alojamentos arrendados, com a renda por m<sup>2</sup> superior ao respetivo percentil a nível nacional, em relação ao número total de alojamentos arrendados, por freguesia, multiplicado por 100.

<sup>17</sup> Número de alojamentos familiares cujo residente era proprietário ou coproprietário da habitação em relação ao número total de alojamentos familiares, por freguesia, multiplicado por 100.

- a proporção de alojamentos familiares cuja aquisição produzia um encargo financeiro mensal igual ou superior a 400€<sup>18</sup>;
- a proporção de indivíduos cujo nível de ensino mais elevado completo corresponde ao Ensino pós-secundário (cursos de especialização tecnológica), ao Bacharelato, à Licenciatura, ao Mestrado ou ao Doutoramento<sup>19</sup>;
- a proporção de indivíduos que pertencia ao grupo socioeconómico “Diretores e quadros dirigentes do Estado e empresas” ou ao grupo “Dirigentes de pequenas empresas e organizações”<sup>20</sup>;

### *3.5. Agregação de freguesias*

A fim de se garantir a robustez dos resultados, foi definido um número mínimo de dados por freguesia. Assim, todas as freguesias que apresentaram um número de registos inferior a 15, foram excluídas da análise. Contudo, visto que estas freguesias totalizavam um número significativo, reduziu-se o valor de referência de 15 para 10. Mesmo assim, mais de 70% das freguesias não foram representadas.

Dado este entrave, apresentou-se como solução a criação de uma nova unidade geográfica que contempla as freguesias com pelo menos 10 registos e grupos de freguesias que somem pelo menos esse mesmo valor. Para esta agregação, utilizou-se como fatores de agrupamento o número de registos dos preços da habitação de 2017, assim como a classificação tipológica das áreas urbanas de 2014 (TIPAU 2014). Esta classificação traduz-se numa nomenclatura territorial que permite agrupar as freguesias em três grandes áreas: “Áreas predominantemente urbanas (APU)”, “Áreas mediantemente urbanas (AMU)” e “Áreas predominantemente rurais (APR)”, respeitando a divisão administrativa da versão da Carta Administrativa Oficial de Portugal de 2013 (CAOP 2013). Esta categorização tem por base critérios como a densidade populacional, o número de

---

<sup>18</sup> Número de alojamentos familiares cuja aquisição produzia um encargo financeiro mensal igual ou superior a 400€, em relação ao número total de alojamentos familiares cuja aquisição produzia um encargo financeiro mensal, por freguesia, multiplicado por 100.

<sup>19</sup> Número de indivíduos cujo nível de ensino mais elevado completo corresponde ao Ensino pós-secundário (cursos de especialização tecnológica), ao Bacharelato, à Licenciatura, ao Mestrado ou ao Doutoramento, em relação ao número total de indivíduos, por freguesia, multiplicado por 100.

<sup>20</sup> Número de indivíduos que pertencia ao grupo socioeconómico “Diretores e quadros dirigentes do Estado e empresas” ou ao grupo “Dirigentes de pequenas empresas e organizações”, em relação ao número total de indivíduos que estava empregado ou desempregado mas já trabalhou, por freguesia, multiplicado por 100.



população residente, o nível de propensão do solo para o processo de urbanização e de edificação e a localização da sede da Camara Municipal. Esta classificação apresenta um segundo nível em que as freguesias são agregadas em sub-regiões respeitando a sua tipologia. Este agrupamento garante que o seu limite é confinado ao limite do município a que as freguesias pertencem. No caso das “áreas urbanas” (sub-região de APU) é ainda assegurada a contiguidade espacial entre as freguesias pertencentes ao mesmo grupo (INE, 2014a). Sendo que alguns municípios são compostos só por freguesias em que cada uma apresenta pelo menos 10 registos, estes não foram incluídos na agregação de freguesias e entraram para a base de dados como freguesias autónomas.

Assumindo que as freguesias de interesse para o estudo são tendencialmente mais urbanas, a agregação foi feita em dois grupos distintos: as áreas urbanas e conjuntamente as sub-regiões de AMU e APR. Deste modo, para o segundo grupo foi necessário garantir a contiguidade entre as áreas agregadas<sup>21</sup>. Aquelas cuja contiguidade não foi verificada, foram divididas em duas ou mais áreas de modo a assegurar este requisito. Depois desta validação, desagregaram-se todas as áreas compostas por freguesias em que cada uma apresentasse o número mínimo de registos. Após esta fase, identificaram-se das restantes áreas, aquelas que totalizavam pelo menos 10 registos e as que, pelo contrário, seriam classificadas como um problema a resolver posteriormente. Como resultado desta agregação, das 3092 freguesias, 639 mantiveram-se autónomas, 2389 foram agregadas em 265 regiões e 64 ficaram por encontrar uma nova solução. O esquema que resume esta agregação pode ser observado no Anexo A, figura 23.

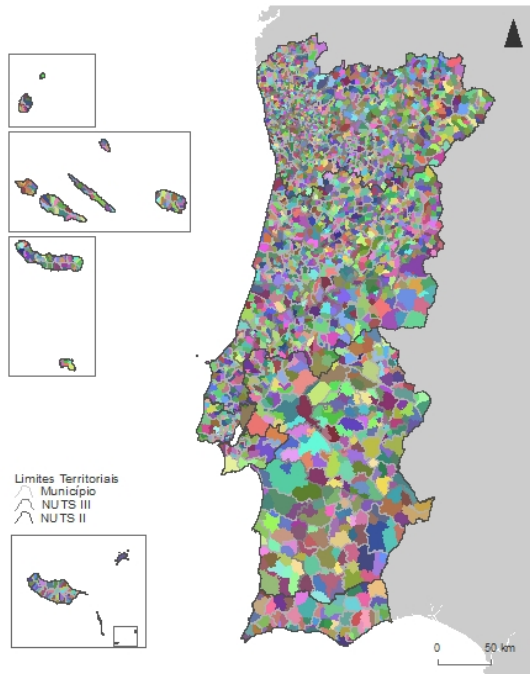
Centrando nos grupos identificados como problemas, foi feita uma análise ad hoc para perceber quais eram as freguesias contíguas que somavam pelo menos 10 registos e aquelas que não se encontrando nesta situação podiam ser agrupadas a outras freguesias já agregadas. Nesta agregação final, foi tido como critérios a contiguidade espacial, o confinamento ao limite do município e a totalização de pelo menos 10 registos. Também se considerou a impossibilidade de agregar estas áreas a freguesias autónomas ou a áreas urbanas, a fim de manter a homogeneidade destas. Assim, foi possível agregar mais 23 freguesias, ficando apenas 41 excluídas da análise.

---

<sup>21</sup> A contiguidade espacial entre as áreas agregadas foi verificada através do *software* ArcGIS.

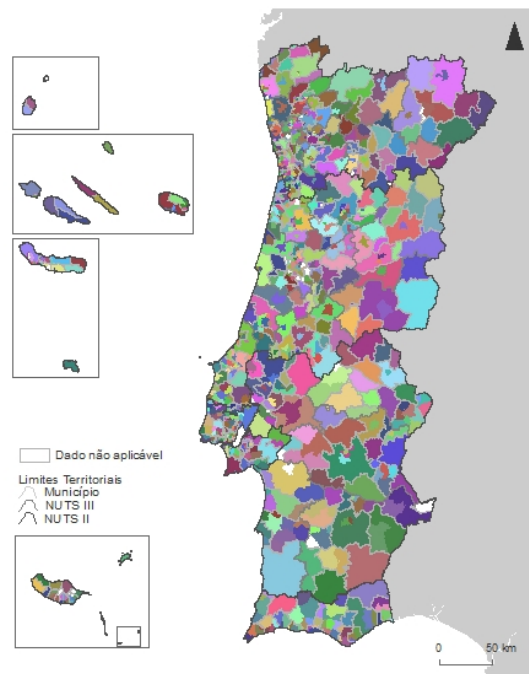
As figuras seguintes representam a Divisão administrativa por freguesia (Figura 1) e a agregação de freguesias acima descrita (Figura 2). Em resultado da sua comparação, pode-se observar uma agregação de freguesias mais acentuada, isto é, a agregação de um maior número de freguesias, no interior relativamente ao litoral.

Figura 1 - Divisão administrativa por freguesia, 2014<sup>22</sup>



Fonte: Carta Administrativa Oficial de Portugal de 2014

Figura 2 - Agregação de freguesias segundo a TIPAU e os preços de habitação, 2014 e 2017



Fonte: TIPAU 2014 e Estatísticas de preços da habitação

Uma vez criada a agregação de freguesias, foram calculados, para esta nova unidade geográfica, as mesmas variáveis já referidas. Deste modo, seguem-se as análises, em paralelo, das freguesias agregadas e das freguesias não agregadas.

### 3.6. Análise Univariada

A análise univariada distingue-se por ser a primeira exploração dos dados a ser realizada, sendo cada variável tratada de forma isolada. Medidas de localização de tendência central e tendência não central e medidas de dispersão permitem perceber como os dados se distribuem e quão dispersos estão. Para esta análise, considerou-se mais relevante as seguintes estatísticas:

<sup>22</sup> Todos os mapas foram criados através do *software* ArcGis.

- Proporção do número de unidades geográficas que apresentem a respetiva variável igual a zero;
- Proporção do número de unidades geográficas que apresentem a respetiva variável sem valor;
- Valor mínimo, máximo e respetiva amplitude;
- Primeiro, segundo e terceiros quartis;
- Valor médio
- Desvio-padrão e coeficiente de variação

De acordo com os resultados, destaca-se a proporção de freguesias que apresentam a respetiva variável sem valor que, como referido no subcapítulo anterior, pode ser um impedimento ao estudo. Neste contexto, salientam-se as variáveis associadas às estatísticas de rendas da habitação que apresentam valores entre os 71,35% e 74,19%, no caso das freguesias não agregadas.

### *3.7. Análise Bivariada*

A análise bivariada permite analisar a relação entre duas variáveis. Sendo a correlação entre duas variáveis uma das formas mais precisas de se medir como elas se relacionam, foram calculadas duas matrizes de correlação (uma para cada unidade geográfica utilizada) das diversas variáveis calculadas. Para tal recorreu-se ao coeficiente de correlação de Pearson, que varia entre -1 e 1. De acordo com as matrizes de correlação (Anexo A, figuras 24 e 25), destacam-se as correlações acima de 0,7 entre as variáveis associadas ao rendimento e as variáveis derivadas dos censos. Também as variáveis do mercado imobiliário (preços e rendas da habitação) apresentam uma forte correlação entre si ( $>0,7$ ). No patamar abaixo ( $>0,3$  e  $<0,7$ ) encontra-se a maioria das correlações, salientando-se, por exemplo, a correlação (acima de 0,5) entre os percentis 80 e 90 do rendimento global e todos os percentis do preço por m<sup>2</sup> dos alojamentos vendidos.

Após a análise uni e bivariada, foram selecionadas as variáveis mais adequadas ao estudo. Por conseguinte, tendo em conta o enfoque do trabalho (as famílias com maior riqueza), utilizaram-se as variáveis referentes aos percentis de valor mais elevado (80 e 90). Salienta-se também que não foram incluídas as variáveis com origem nas bases de dados dos censos e das rendas nos últimos modelos da análise ensaiados e aqui apresentados,

devido, essencialmente, ao reduzido número de observações e à distância temporal no caso das variáveis censitárias. No que diz respeito às variáveis relativas ao rendimento dos agregados familiares, selecionou-se para os modelos finais, aquelas que se referem ao rendimento global do agregado por este identificar o rendimento auferido e assim se aproximar mais do objetivo deste estudo. Deste modo, tem-se:

<b>per80_rendimento_global16</b>	percentil 80 do rendimento global dos agregados familiares, em 2016, por freguesia/região.
<b>per90_rendimento_global16</b>	percentil 90 do rendimento global dos agregados familiares, em 2016, por freguesia/região.
<b>per80_PT_rendimento_global16</b>	proporção de agregados familiares com o rendimento global superior ao percentil 80 a nível nacional (23725.36€), em 2016, por freguesia/região.
<b>per90_PT_rendimento_global16</b>	proporção de agregados familiares com o rendimento global superior ao percentil 90 a nível nacional (35904.45€), em 2016, por freguesia/região.

Relativamente à base de dados dos preços da habitação, focou-se nas variáveis associadas ao preço por metro quadrado por englobarem a maior parte da informação necessária, para os níveis de percentis mais elevados e para o ano mais recente (2017), tendo-se:

<b>per80Vm2_phab17</b>	percentil 80 do valor por m <sup>2</sup> da venda de habitações, em 2017, por freguesia/região.
<b>per90Vm2_phab17</b>	percentil 90 do valor por m <sup>2</sup> da venda de habitações, em 2017, por freguesia/região.
<b>per80Vm2_PT_phab17</b>	proporção de habitações vendidas com valor por m <sup>2</sup> superior ao percentil 80 a nível nacional (1597.39€), em 2017, por freguesia/região.

**per90Vm2\_PT\_phab17** proporção de habitações vendidas com valor por m<sup>2</sup> superior ao percentil 90 a nível nacional (2187.50€), em 2017, por freguesia/região.

### 3.8. Análise fatorial

A análise fatorial é uma análise exploratória de dados (desenvolvida por Charles Spearman, em 1904) que visa analisar um conjunto de variáveis inter-relacionadas a fim de encontrar o(s) fator(es) comum(ns) latente(s) (características em comum, normalmente não observáveis) e as relações estruturais entre eles, através das correlações observadas entre as variáveis. O modelo de m fatores pode ser representado como:

$$\begin{aligned}x_1 &= \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \cdots + \lambda_{1m}\xi_m + \epsilon_1 \\x_2 &= \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \cdots + \lambda_{2m}\xi_m + \epsilon_2 \\&\vdots \\x_p &= \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \cdots + \lambda_{pm}\xi_m + \epsilon_p\end{aligned}$$

Onde  $\xi_m$  corresponde aos fatores comuns,  $\epsilon_p$  corresponde aos fatores únicos (específicos de cada variável, também identificados como o erro) e  $\lambda_{pm}$  representa o peso da variável p no fator comum m (designados por *factor loadings*).

Segundo Sharma (1996), os métodos de extração de fatores mais utilizados são o Método das Componentes Principais e o Método da Fatorização do Eixo Principal. Estas técnicas distinguem-se, essencialmente, por a primeira assumir que cada variável é constituída apenas por fatores comuns e o número de fatores ser igual ao número de componentes ( $m = p$ ) enquanto na segunda técnica cada variável ser composta por fatores comuns e o respetivo fator único.

O modelo de análise subjacente a este trabalho corresponde a uma análise fatorial em componentes principais. Na fase inicial em que, segundo esta técnica, é assumido que as estimativas iniciais das comunalidades<sup>23</sup> para todas as variáveis são iguais a um, a matriz de correlação com as comunalidades estimadas na diagonal é submetida à análise em componentes principais. A fase seguinte consiste na extração fatorial que cria combinações lineares não correlacionadas das variáveis iniciais, possível de ser realizada

<sup>23</sup> A comunalidade corresponde à proporção da variância de cada variável explicada pelos fatores comuns. (Marôco, 2010, p. 490.)

através de vários critérios. Como critério de extração adotou-se aquele cujos valores próprios são superiores a 1 (Critério de Kaiser). O primeiro fator contém a variância máxima extraída enquanto os seguintes apresentam uma variância sucessivamente mais reduzida. Assim, é garantido que as componentes extraídas apresentam uma variabilidade superior à variância individual de cada variável de base. Na fase seguinte, procedeu-se à rotação dos fatores extraídos através do método de rotação ortogonal varimax cujo objetivo é obter uma estrutura fatorial na qual cada variável esteja fortemente associada a apenas um fator. (Sharma, 1996).

### *3.9. Análise de clusters*

Esta secção é baseada em Marôco (2010). A análise de clusters é uma técnica exploratória de análise multivariada que permite reunir sujeitos ou variáveis, em grupos homogêneos, relacionados entre si através de uma ou mais características. Assim, cada observação pertencente a uma classe/cluster terá características semelhantes a outras observações pertencentes ao mesmo grupo e diferentes das restantes pertencentes a outros clusters. Esta análise pode ser aplicada sobre um conjunto de variáveis de base ou sobre as componentes extraídas de uma análise fatorial. Os estudos publicados referentes à tipologia socioeconómica das áreas metropolitanas de Lisboa e Porto 2001 e 2011 (INE, 2004a; INE, 2004b; INE, 2014b) são exemplos da utilização das componentes resultantes da análise fatorial, para a análise de clusters.

Inicialmente, aquando da classificação dos dados iniciais, nem o número de clusters nem os membros do grupo são conhecidos (Pestana e Gageiro, 2000, p.429-446). Por conseguinte, para a identificação dos agrupamentos é necessário medir a semelhança entre as características. Quando as variáveis introduzidas são duas ou três, é possível que os grupos homogêneos sejam facilmente identificáveis e a solução se possa interpretar graficamente. Porém, dado um número de variáveis superior a este, é necessário recorrer a outros métodos como as medidas de semelhança (proximidade) ou dissemelhança (distância) entre dois casos ou posteriormente entre clusters (Marôco, 2010, p.549).

Marôco (2010, p.549-552) apresenta como medidas mais frequentes a Distância Euclidiana, a Distância de Minkowski, a Distância de Mahalanobis, a Medida de Semelhança do Coseno, o Coeficiente de Jaccard, de Russel & Rão e medidas de associação binárias e por fim, as medidas de semelhança para variáveis.

Uma vez calculadas as medidas de dissimilaridade, segue-se o agrupamento dos sujeitos ou variáveis em clusters. Pretende-se que estas medidas sejam o menor possível dentro dos grupos e o maior possível entre eles. Para tal, existem duas técnicas: O Agrupamento hierárquico de Clusters (*Hierarchical*) e o Agrupamento Não-Hierárquico de Clusters (*Partitional*).

O primeiro método traduz-se em sucessivos passos de aglomeração (aglomerantes) ou de divisão (divisivos) em que, de acordo com as proximidades entre os sujeitos, estes vão sendo aglomerados ou o cluster inicial que contém todos os sujeitos vai sendo dividido em subgrupos, respetivamente. Relativamente aos métodos aglomerantes, o algoritmo procura na matriz de dissimilaridade os pares de sujeitos e/ou clusters de menor distância e de seguida reúne-os, sendo que na última iteração existirá um único cluster com todos os casos. Após o primeiro passo, a união entre dois objetos, surge a necessidade de calcular a medida de proximidade entre os elementos pertencentes ao cluster inicial e os restantes elementos. Segundo Marôco (2010, p.554-555), esta pode ser obtida através de um dos seguintes critérios:

- **Menor Distância (Single Linkage):** a menor distância entre cada elemento constituinte do cluster e os restantes;
- **Maior distância (Complete Linkage):** a maior distância entre cada elemento constituinte do cluster e os restantes;
- **Distância média entre clusters:** a média das distâncias entre cada elemento pertencente ao cluster e os restantes objetos;
- **Distância média dentro dos clusters:** a menor soma de quadrados dos erros (i.e. variabilidade dentro dos Clusters) uma vez os clusters unidos. Esta medida é menos sensível a outliers;
- **Distância com a Mediana:** a mediana das distâncias de cada objeto constituinte do cluster e os restantes elementos;
- **Método do Centróide:** distância entre o ponto médio (cujas coordenadas são a média dos objetos que pertencem ao cluster) e os valores dos restantes elementos;
- **Método de Ward:** a menor soma dos quadrados dos erros uma vez os clusters unidos, sendo cada um representado pelo seu centróide;

O melhor método a ser utilizado não é definido à partida, depende da natureza dos dados e dos resultados pretendidos. Por isso, é recomendado utilizar-se vários em simultâneo. Se todos produzirem soluções interpretáveis semelhantes, então conclui-se que a matriz dos dados apresenta agrupamentos “naturais” e não “artificiais”.

No que diz respeito à escolha do número ideal de Clusters, esta pode ser feita através da visualização de todas as etapas de agregação (ou divisão) num dendograma. Neste gráfico, o conjunto de clusters está organizado como uma árvore. Cada nó (cluster) une os respetivos subconjuntos e a raiz da árvore corresponde ao Cluster que contém todos os objetos. Além disto, de acordo com Tan (2018), o Agrupamento Hierárquico pode ser visto como uma sequência de Agrupamentos Não-Hierárquicos que se podem obter uma vez cortada a árvore em vários níveis.

O Agrupamento Não-Hierárquico de clusters consiste na junção de itens (e não variáveis) num conjunto de Clusters cujo número é definido inicialmente ( $k$ ). Marôco (2010) aponta como principais vantagens, face aos métodos hierárquicos, a facilidade com que são aplicados às matrizes de dados muito grandes e a capacidade de reagrupar elementos que já pertenceram a outro cluster diferente do inicial. Esta situação não é possível nos métodos hierárquicos uma vez que a inclusão de um sujeito num cluster é fixa. Consequentemente, a classificação errada de um sujeito é muito improvável nesta técnica.

Também o Agrupamento Não-Hierárquico de clusters apresenta diversos métodos. Estes diferem essencialmente na primeira agregação e no cálculo das novas distâncias entre os centróides dos clusters e os restantes sujeitos. O método das  $k$ -médias é o mais utilizado e dos mais simples algoritmos de aprendizagem não supervisionados. Caracteriza-se por cada observação ser agrupada num cluster tendo em conta o protótipo (centróide) mais próximo e a minimização da soma do quadrado dos desvios intra-cluster. O algoritmo termina quando não existir variação significativa da distância mínima ou quando o número máximo de interações ou o critério de convergência forem atingidos. Note-se que, assim que a medida de proximidade e a função objetivo forem definidas, é possível definir  $k$  matematicamente. Uma das desvantagens que o  $k$ -médias apresenta é a possibilidade da existência de clusters vazios. Neste caso, é necessário obter um centróide de substituição. Este pode ser o ponto que se encontra mais afastado de qualquer centróide ou um ponto do cluster com a maior variabilidade. Em qualquer caso, a introdução do



centróide de substituição irá reduzir a soma total dos erros quadrados (*SSE – Sum of Squared Errors*). Outras ações podem diminuir esta variabilidade: a eliminação de um outlier ou de pequenos grupos provavelmente constituídos por outliers. Assim, se o *k* definido inicialmente for elevado e posteriormente se aplicar este processo de eliminação de outliers/clusters, provavelmente obter-se-á um reduzido SSE e um reduzido número de clusters, visto que o algoritmo *k*-médias apresenta uma tendência para convergir para um mínimo local. Existem ainda duas técnicas que permitem a redução simultânea do número de clusters e do SSE total: eliminando o centróide e alocando os pontos do respetivo cluster aos clusters mais próximos; ou unindo os clusters com centróides mais próximos.

No presente exercício considerou-se o método hierárquico de agregação, uma vez que não se pressupõe à partida o número de clusters (grupos) que se encontram na matriz de dados. O método de agregação utilizado consiste no método Ward que, segundo Sharma (1996), cria clusters maximizando a homogeneidade intra-clusters através da minimização da soma dos erros quadrados (SSE) dentro do grupo (cluster), avaliada a cada etapa de agregação. Como medida de dissimilaridade empregou-se o quadrado da distância euclidiana definida como:

$$D_{ij}^2 = \sum_{k=1}^p (x_{ik} - x_{jk})^2$$

Onde  $D_{ij}^2$  é o quadrado da distância euclidiana entre os casos *i* e *j*,  $x_{ik}$  é o valor da *k*-ésima variável para o *i*-ésimo caso,  $x_{jk}$  é o valor da *k*-ésima variável para o *j*-ésimo caso e *p* é o número total de variáveis.

Todas as variáveis introduzidas foram estandardizadas pelo método do *z*-scores que garante que a distribuição apresenta uma média igual a zero e desvio-padrão igual a um.

Para uma análise da evolução e consistência do grupo de interesse, os resultados da análise de clusters foram calculados, em cada modelo, para os casos de três, quatro e cinco clusters.

#### 4.Resultados

Em resultado dos diversos estudos realizados, selecionou-se quatro modelos que melhor se enquadram com o objetivo deste trabalho. Nos dois primeiros estudos foram aplicadas análises de clusters para classificar as freguesias/regiões em termos de riqueza, diferenciando-se apenas o valor do percentil das variáveis introduzidas. No primeiro as variáveis referem-se ao percentil 80 e no segundo ao percentil 90. Uma vez que os resultados das regiões mais ricas apontavam para áreas fortemente associadas ao turismo cuja característica faz parte da variável referente aos preços de habitação, realizou-se uma análise fatorial. Com esta análise pretendeu-se expurgar a componente do turismo associada à variável dos preços de habitação, através da adição de uma nova variável - o rácio de alojamentos clássicos classificados como alojamento local, em 2016, por freguesia/região (prop\_AL). Esta análise teve por base as variáveis utilizadas nos dois estudos enunciados, surgindo assim o terceiro e quarto modelos. Em cada análise, foram extraídos dois fatores. O primeiro apresenta uma maior correlação com as variáveis relativas ao rendimento (superior a 95%) e negativa em relação à variável alojamento local, enquanto que o segundo mostra uma maior correlação (superior a 85%) com a variável associada ao turismo (alojamento local). Assim, sendo o primeiro fator interpretado como aquele que traduz o nível de riqueza associada a cada freguesia/região, foi submetida uma análise de clusters a esse fator.

Nos gráficos que se seguem, pode-se observar, para cada estudo, o comportamento das regiões de acordo com cada uma das quatro variáveis introduzidas e o coeficiente de variação a elas associado. Salienta-se que esta informação é apresentada segundo o agrupamento de regiões em três clusters, dado que o número de regiões correspondente se adequa mais à futura seleção de amostra para o ISFF e porque permite uma melhor visualização dos resultados. A azul estão identificadas as regiões pertencentes ao cluster classificado como o que apresenta um maior nível de riqueza e a verde, as restantes regiões não pertencentes a este grupo. Tem-se ainda, como referência (a laranja), a média do cluster segundo a variável em análise. Esta distinção permite identificar se as regiões com valores mais elevados de percentis pertencem ao cluster de interesse e avaliar a homogeneidade intrarregiões. Além disso, observando o gráfico como um todo, pode-se avaliar a homogeneidade do referido cluster. Neste capítulo apenas serão apresentados os resultados por região, ou seja, para as freguesias agregadas. Os mapas temáticos

referentes a cada um destes estudos, assim como os resultados por freguesia podem ser consultados no Anexo B.

#### 4.1. Estudo 1

Para o primeiro modelo, foram introduzidas quatro variáveis: **per80\_rendimento\_global16**, **per80\_PT\_rendimento\_global16**, **per80Vm2\_phab17** e **per80Vm2\_PT\_phab17**. Com base na figura 3, que mostra o agrupamento de elementos (regiões) para as divisões em 3, 4 e 5 clusters, verifica-se que o grupo de interesse (destacado a azul escuro) manteve-se consistente, sendo composto por cinquenta e quatro regiões que correspondem a cinquenta e quatro freguesias.

Figura 3 – Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 1.

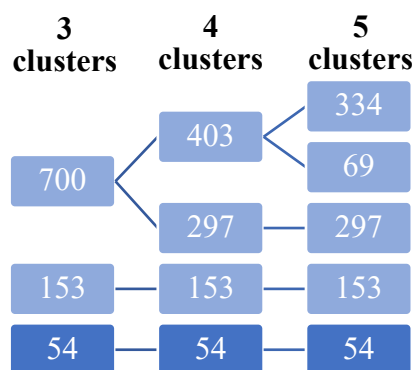


Figura 4 - Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 1, por região, 2016

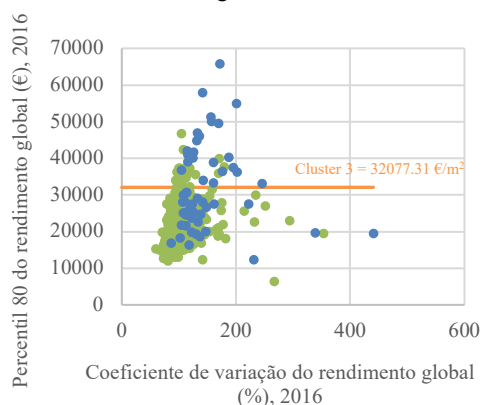
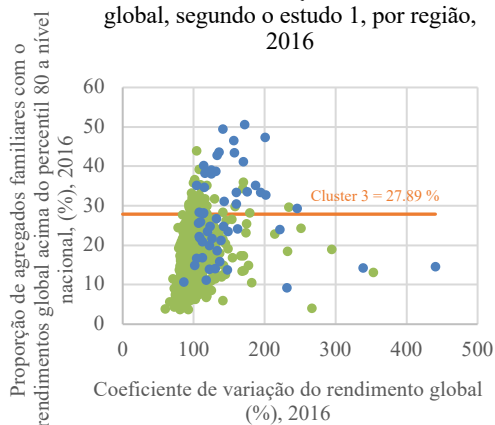


Figura 5 - Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 1, por região, 2016



Fonte: Informação associada às notas de liquidação do IRS, 2016

Figura 6 - Percentil 80 dos preços por m<sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por região, 2017

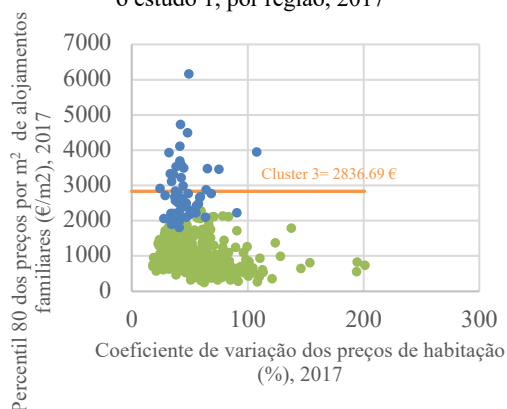
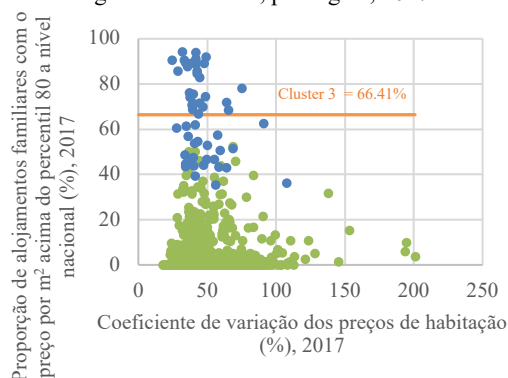


Figura 7 - Proporção de alojamentos familiares com o preço por m<sup>2</sup> acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por região, 2017



Fonte: Estatísticas dos preços da habitação, 2017

#### 4.2. Estudo 2

No segundo estudo, utilizaram-se as seguintes variáveis: **per90\_rendimento\_global16**, **per90\_PT\_rendimento\_global16**, **per90Vm2\_phab17** e **per90Vm2\_PT\_phab17**. Também para este modelo, o grupo associado a maior riqueza manteve-se consistente, mas neste caso com apenas vinte elementos (freguesias) - Figura 8.

Figura 8 – Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 2.

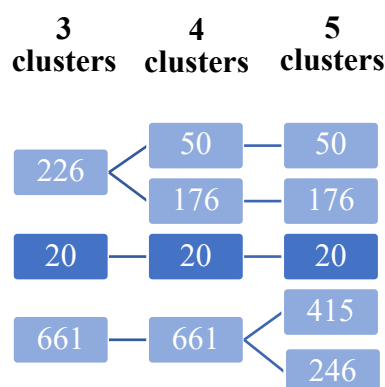


Figura 9 - Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 2, por região, 2016

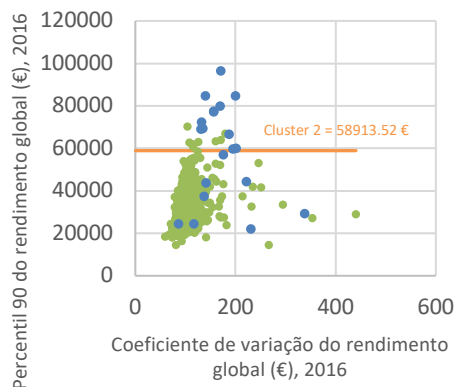
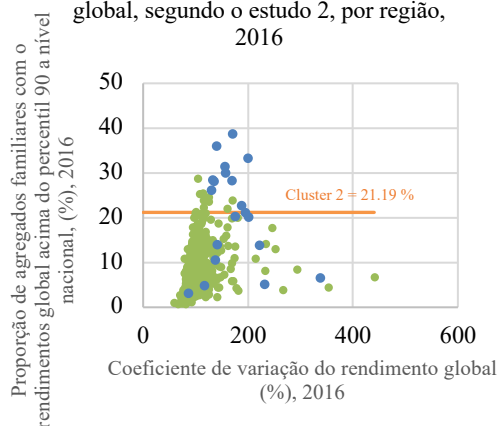


Figura 10 - Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 2, por região, 2016



Fonte: Informação associada às notas de liquidação do IRS, 2016

Figura 11 - Percentil 90 dos preços por m<sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por região, 2017

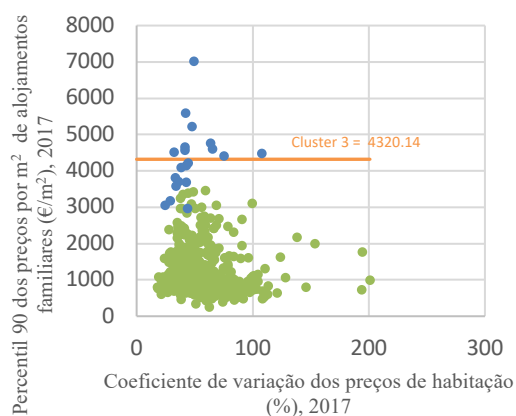
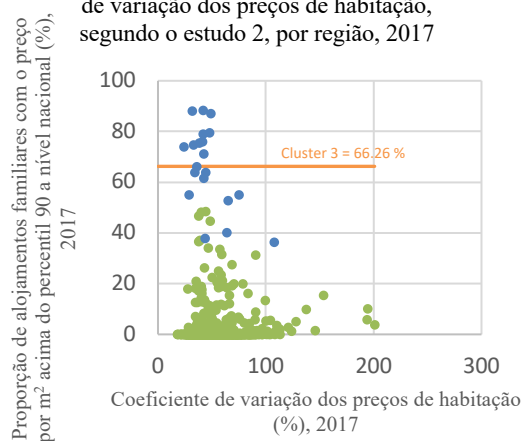


Figura 12 - Proporção de alojamentos familiares com o preço por m<sup>2</sup> acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por região, 2017



Fonte: Estatísticas dos preços da habitação, 2017

#### 4.3. Estudo 3

Neste modelo, comparando ao estudo 1, foi adicionada uma variável, tendo sido incluídas as cinco variáveis: **per80\_rendimento\_global16**, **per80\_PT\_rendimento\_global16**, **per80Vm2\_phab17**, **per80Vm2\_PT\_phab17** e **prop\_AL**. Após a análise fatorial, os dois fatores retidos neste modelo revelam uma capacidade explicativa de 90,76% da variância total das cinco variáveis, sendo que o primeiro fator explica isoladamente (após rotação) 45,62% da variância total das variáveis e o segundo 45,14%. Da análise de cluster com o primeiro fator, resultaram 147 regiões do grupo com maior média (nas variáveis), no cenário de três clusters, correspondendo a 166 freguesias (Anexo B, tabela

II). No caso da divisão em quatro e cinco clusters, este grupo (147) é separado em dois, sendo o de 15 regiões aquele que apresenta uma média superior (Figura 13).

Figura 13 – Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 3.

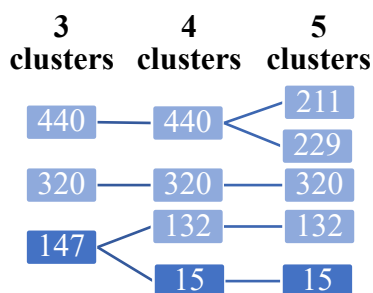


Figura 14 - Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 3, por região, 2016

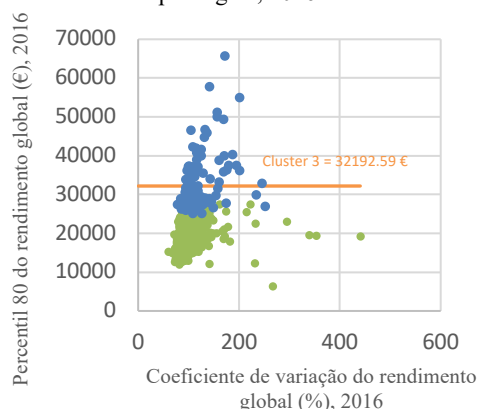
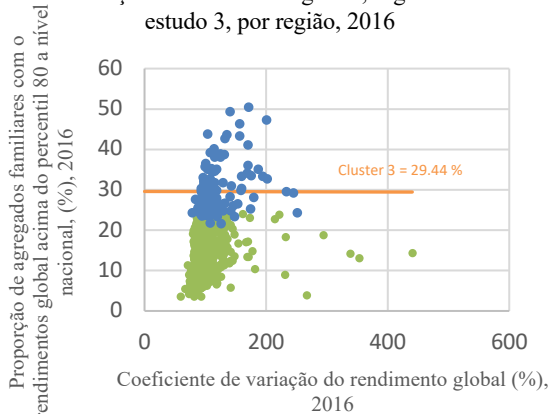


Figura 15 - Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 3, por região, 2016



Fonte: Informação associada às notas de liquidação do IRS, 2016

Figura 16 - Percentil 80 dos preços por m² de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 3, por região, 2017

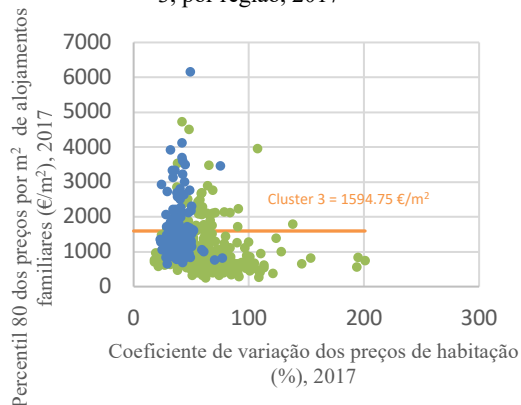
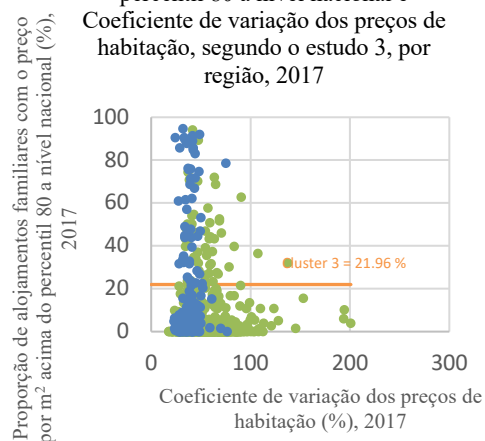


Figura 17 - Proporção de alojamentos familiares com o preço por m² acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 3, por região, 2017



Fonte: Estatísticas dos preços da habitação, 2017

#### 4.4. Estudo 4

As variáveis do estudo 4 correspondem às do estudo 2 com adição da variável relativa ao alojamento local. Assim, foram utilizadas as variáveis: **per90\_rendimento\_global16**, **per90\_PT\_rendimento\_global16**, **per90Vm2\_phab17**, **per90Vm2\_PT\_phab17** e **prop\_AL**. Com base na análise fatorial, os dois fatores retidos neste modelo revelam uma capacidade explicativa de 89,52% da variância total das cinco variáveis, sendo que o primeiro fator explica isoladamente (após rotação) 46,31% da variância total das variáveis e o segundo explica 43,20%. De acordo com a análise de cluster, o grupo com maior média para os casos de três e quatro clusters apresenta 51 regiões. No caso da divisão em cinco clusters, este grupo é dividido em dois (Figura 18), sendo o de 10 freguesias aquele que apresenta uma média superior.

Figura 18 – Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 4

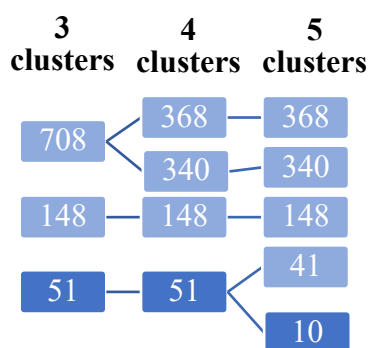


Figura 19 - Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 4, por região, 2016

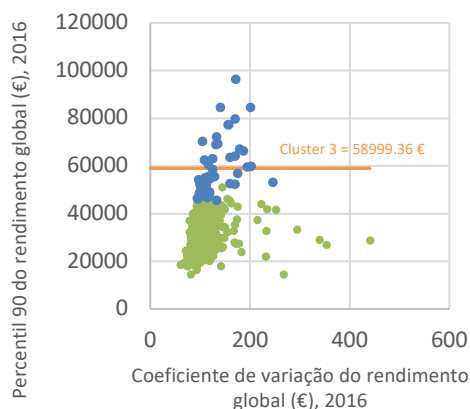
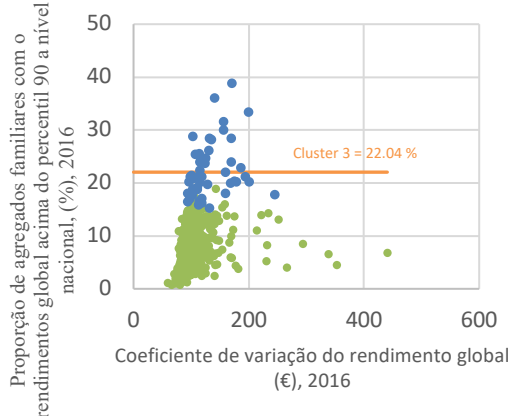


Figura 20 - Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 4, por região, 2016



Fonte: Informação associada às notas de liquidação do IRS, 2016

Figura 21 - Percentil 90 dos preços por m<sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 4, por região, 2017

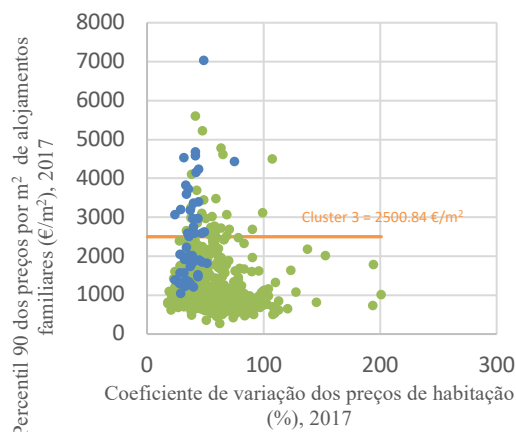
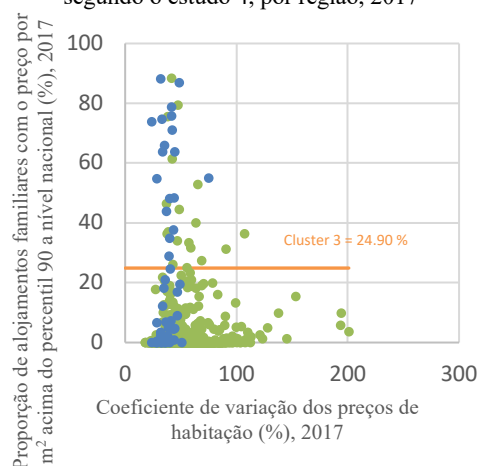


Figura 22 - Proporção de alojamentos familiares com o preço por m<sup>2</sup> acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 4, por região, 2017



Fonte: Estatísticas dos preços da habitação, 2017

#### 4.5. *Análise de Resultados*

Analizados os quatro modelos expostos, conclui-se que os estudos 1 e 2 apresentam, durante o agrupamento de clusters, uma maior consistência do grupo de regiões com níveis potencialmente mais elevados de rendimento e património. Observou-se ainda que neste grupo e relativamente às variáveis dos preços da habitação, os seus valores são os mais elevados, enquanto que os valores das variáveis do rendimento global surgem mais dispersos. Esta situação inverte-se nos estudos 3 e 4 em que as regiões que compõem este grupo apresentam valores mais elevados para as variáveis do rendimento global do que para as variáveis dos preços da habitação. Assim, é de notar que a homogeneidade intracuster varia consoante a variável observada. Com base nos mapas temáticos apresentados no anexo B (figuras 26 a 29), verifica-se que as regiões do cluster com um nível de riqueza mais elevado correspondem a freguesias autónomas, com exceção do observado no estudo 3 que apresenta quatro regiões que correspondem a freguesias agregadas. Este fato sugere que as freguesias onde o mercado imobiliário é mais dinâmico, são também as freguesias mais prováveis a apresentarem um maior nível de riqueza. Estas áreas podem ser também caracterizadas pela residência de um elevado número de agregados familiares com altos valores de rendimento global declarado.



## 5. Conclusão

Com base nos estudos apresentados, verificaram-se algumas diferenças na identificação dos territórios com níveis de rendimento e património mais elevados. Apesar destas variações, concluiu-se que a presença de treze freguesias (pertencentes aos municípios do Porto, Cascais e Lisboa) é transversal a todos os estudos. Tudo indica que o estudo 3 é aquele que mais se enquadra na seleção de freguesias. Esta escolha justifica-se pela utilização do percentil 80 (que abrange uma maior percentagem de riqueza - 20%) e pelo cluster com um nível de riqueza mais elevado ser composto por um maior número de regiões (147). Estas características possibilitarão uma maior garantia da existência de alojamentos disponíveis para a seleção da amostra. Para além disso, este estudo exclui a componente do turismo que contribuía para que algumas freguesias fossem classificadas erradamente como aquelas onde residem as famílias com um maior nível de riqueza.

No decorrer deste trabalho, surgiram algumas limitações e obstáculos que impossibilitaram a que se chegasse a um resultado mais preciso e desejado. Contudo, tentou-se contornar as situações de modo a obter o melhor resultado. Os mais relevantes residiram na inacessibilidade à informação referente ao património das famílias, à identificação da composição dos agregados familiares (no âmbito do rendimento declarado), assim como ao reduzido número de registos relativos ao mercado imobiliário, impossibilitando neste caso uma cobertura total do território nacional ao nível da freguesia. Tentando ultrapassar esta situação, apostou-se a agregação de algumas freguesias. Não obstante a todas estas dificuldades, considera-se que os resultados obtidos constituem um contributo positivo para o INE.

Futuramente seria interessante utilizar informação sobre o património das famílias, nomeadamente ativos não financeiros, assim como informação sobre a natureza dos rendimentos declarados. Em relação ao reduzido número de alguns dados, existe a expectativa de, em breve, se poder relacionar a informação das Estatísticas dos preços e das rendas da habitação com informação obtida pelos Censos, criando uma base de dados atualizada e abrangente a todo o território nacional. Por fim e de acordo com a 1ª Lei da Geografia (“tudo está relacionado com tudo, mas o que se encontra mais próximo está mais relacionado do que o que se encontra mais distante” – Tobler, 1970), sugere-se uma análise espacial de clusters para possíveis resultados mais ajustados.

### Referências Bibliográficas

Banco de Portugal e Instituto Nacional de Estatística (2012). Inquérito à Situação Financeira das Famílias - 2010.

Bricker, B. e Henriques, A. e Moore, K. (2017). Updates to the Sampling of Wealthy Families in the Survey of Consumer Finances. Finance and Economics Discussion Series 2017-114, Board of Governors of the Federal Reserve System (US). Disponível em: <https://www.federalreserve.gov/econres/feds/files/2017114pap.pdf>.

Costa, S. e Farinha, L. (2012). Household Finance and Consumption Survey: methodology and main results. Occasional Paper No.1.

European Central Bank (2013). The Eurosystem Household Finance and Consumption Survey. Methodological report for the first Wave. Statistics Paper Series No I. Disponível em: <https://www.ecb.europa.eu/pub/pdf/other/ecbsp1en.pdf>.

European Central Bank (2016). The Household Finance and Consumption Survey: methodological report for the second wave. Statistics Paper Series No17. Disponível em: <https://www.ecb.europa.eu/pub/pdf/scpsps/ecbsp17.en.pdf>.

Federal Reserve Bulletin (2014). Changes in U.S. Family Finances from 2010 to 2013: Evidence from the Survey of Consumer Finances. Vol 100, No 4. Disponível em: <https://www.federalreserve.gov/pubs/bulletin/2014/pdf/scf14.pdf>.

Instituto Nacional de Estatística (2004a). Tipologia Sócio-Económica da Área Metropolitana de Lisboa 2001. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2004b). Tipologia Sócio-económica da Área Metropolitana do Porto 2001. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2010). Censos 2011. XV Recenseamento Geral da População. V Recenseamento Geral da Habitação. Documento Metodológico. Versão 2.0. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2013). Inquérito à Situação Financeira das Famílias, Documento Metodológico Versão 1.1. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2014a). Tipologia de áreas urbanas 2014. Relatório técnico. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2014b). Tipologia socioeconómica das áreas metropolitanas de Lisboa e Porto 2011. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2016). Inquérito à Situação Financeira das Famílias, Documento Metodológico Versão 1.1. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2017). Inquérito à Situação Financeira das Famílias, Documento Metodológico Versão 1.2. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2018). Documento metodológico. Estatísticas de Preços da Habitação ao nível local. Versão 1.0. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2019a). Documento Metodológico. Estatísticas de Rendas da Habitação ao nível local. Versão 2.0. Lisboa: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (2019b). Estatísticas do Rendimento ao nível local 2017. Lisboa: Instituto Nacional de Estatística.

Kennickell, A. (2007). The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances. Disponível em: <https://www.federalreserve.gov/econresdata/scf/files/isi2007.pdf>.

Lohr, Sharon L. (2010). Sampling: Design and analysis, 2ª Ed. USA: Cengage Learning.

Marôco, J. (2010). Análise Estatística com PASW Statistics. Lisboa: Ed.Sílabo.

Pestana, M. e Gageiro, J. (2000). Análise de Dados para Ciências Sociais: A Complementaridade do SPSS. 2ª Ed. Lisboa: Edições Sílabo.

Sharma, S. (1996). Applied Multivariate Techniques. 1ªEd. USA: John Wiley & Sons.

Tan, P. e Steinbach, M. e Karpatne, A. e Kumar, V. (2018). Introduction to Data Mining. 2ªEd. Pearson. Disponível em: <https://www-users.cs.umn.edu/~kumar001/dmbook/ch8.pdf>.

## Anexos

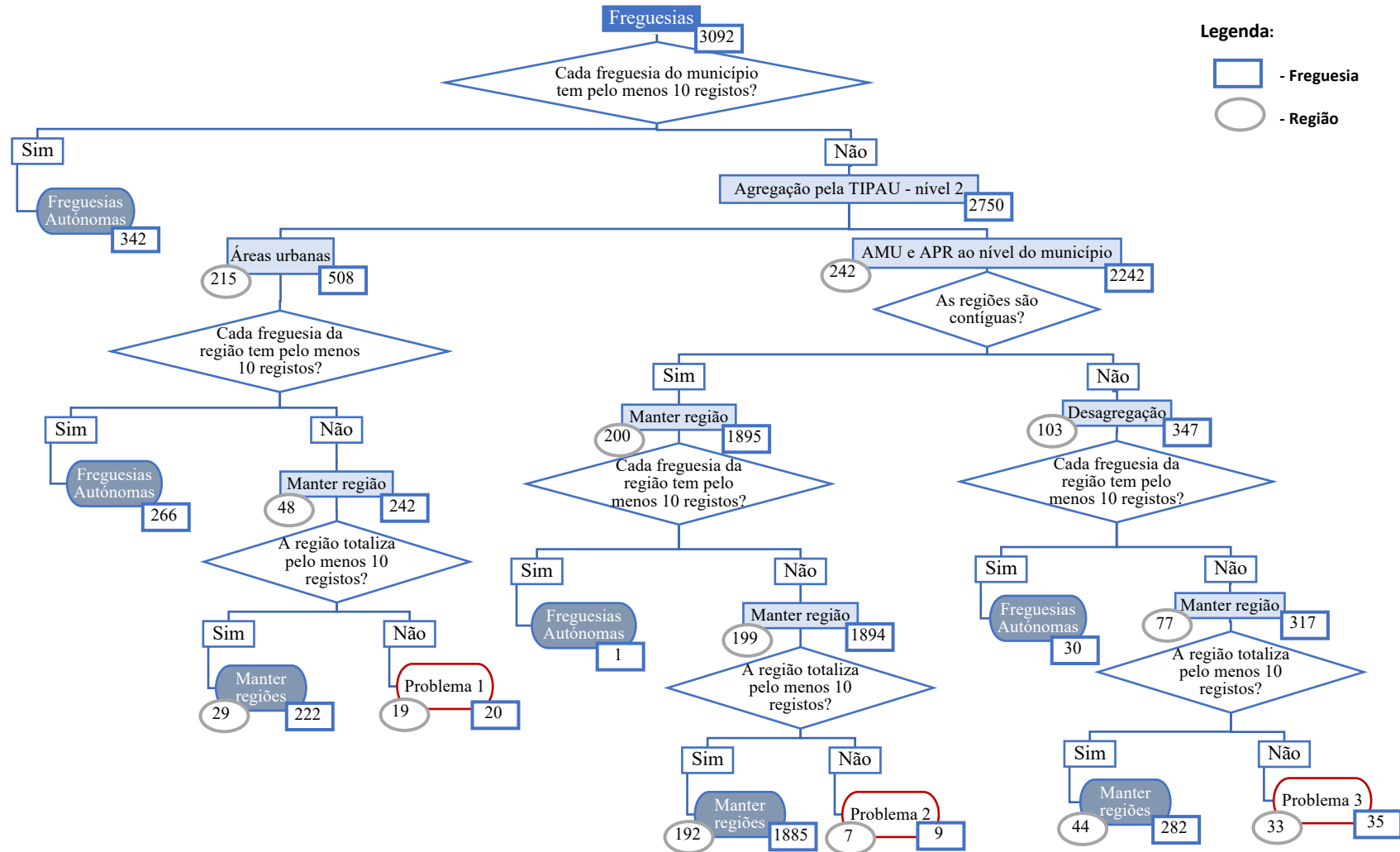
## Anexo A

Tabela I – Distribuição da segunda subamostra de acordo com a zona geográfica e respetiva classe de área útil.

Zona geográfica	Classes de área útil (alojamento)		
	$\geq 100 \text{ m}^2$	$\geq 120 \text{ m}^2$	$\geq 150 \text{ m}^2$
Norte exceto Porto			x
Porto		x	
Centro			x
Grande Lisboa	x		
Península de Setúbal		x	
Alentejo			x
Algarve		x	
Açores			x
Madeira		x	

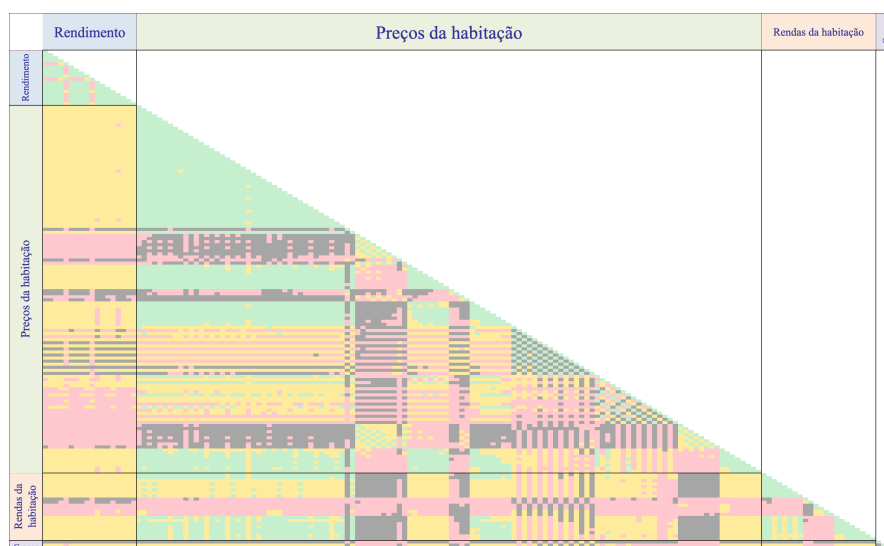
Fonte: Inquérito à Situação Financeira das Famílias, Documento Metodológico, 2013 (INE, 2013)

Figura 23 – Agregação de freguesias



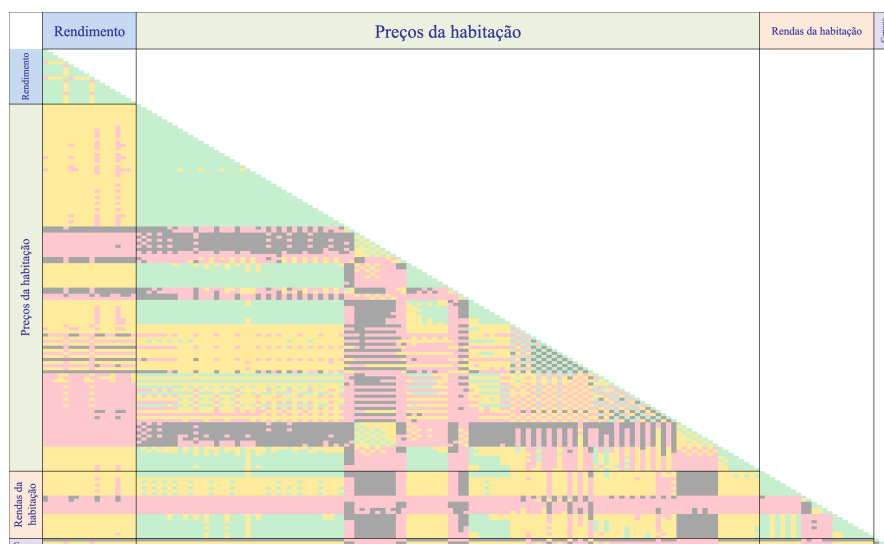
Dado o elevado número de variáveis, optou-se por apresentar de forma genérica os diferentes níveis de correlação entre as variáveis, na matriz de correlação. A cada nível corresponde uma cor: verde a uma correlação acima de 0,7, amarelo entre 0,3 e 0,7, vermelho entre 0 e 0,3 e a cinzento para valores negativos. As matrizes encontram-se divididas pela seguinte ordem de fontes das variáveis: rendimento, preços da habitação, rendas da habitação e censos.

Figura 24 – Matriz de correlação das variáveis por região



Fonte: Informação associada à nota de liquidação do IRS (2016), Estatísticas dos preços da habitação (2017), Estatísticas das rendas da habitação (2017) e Censos (2011).

Figura 25 – Matriz de correlação das variáveis por freguesia



Fonte: Informação associada à nota de liquidação do IRS (2016), Estatísticas dos preços da habitação (2017), Estatísticas das rendas da habitação (2017) e Censos (2011).

Anexo B: Mapas temáticos da análise de clusters por região, para os estudos 1, 2, 3 e 4  
Na legenda das cores de cada figura, lê-se a identificação dos grupos a que pertencem as regiões, no cenário de três clusters. A informação sobre o número de freguesias que corresponde a cada grupo é também apresentada nas figuras.

Figura 26 – Classificação das freguesias para três clusters segundo o Estudo 1

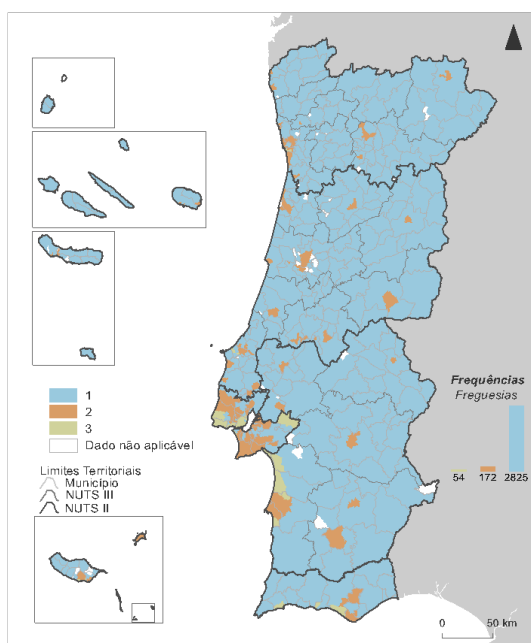
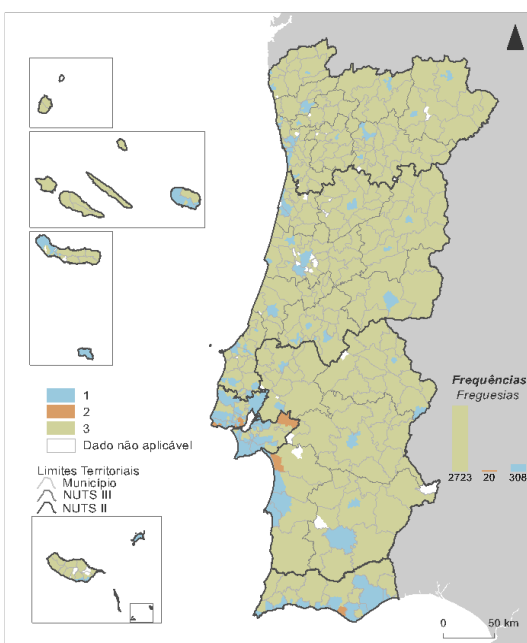


Figura 27 – Classificação das freguesias para três clusters segundo o Estudo 2



Fonte: Informação associada à nota de liquidação do IRS e Estatísticas dos preços de habitação, 2016 e 2017

Figura 28 – Classificação das freguesias para três clusters segundo o Estudo 3

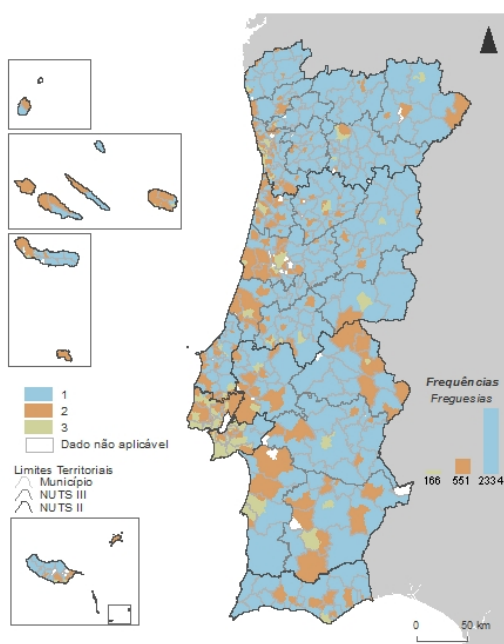
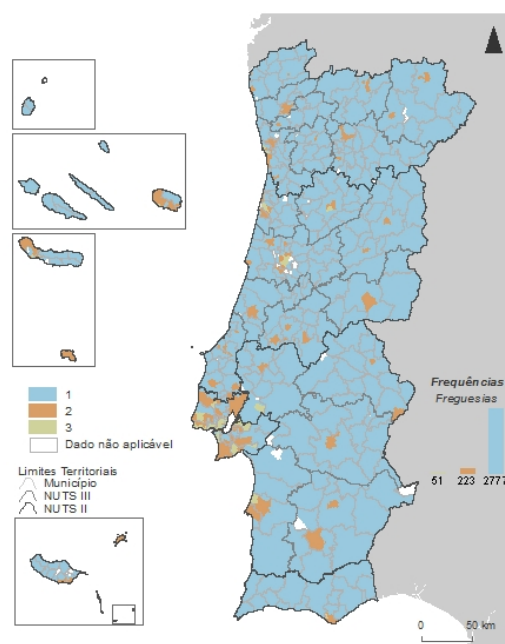


Figura 29 – Classificação das freguesias para três clusters segundo o Estudo 4



Fonte: Informação associada à nota de liquidação do IRS e Estatísticas dos preços de habitação, 2016 e 2017

Tabela II – Códigos e designações da divisão administrativa para o 2º e 3º níveis (municípios e freguesias, respetivamente) de acordo com o cluster com um nível mais elevado de riqueza do estudo 3.

Código de município	Designação do município	Código de freguesia	Freguesia
0105	Aveiro	010501	Aradas
0105	Aveiro	010505	Esgueira
0105	Aveiro	010510	São Bernardo
0105	Aveiro	010513	Santa Joana
0105	Aveiro	010517	União das freguesias de Glória e Vera Cruz
0107	Espinho	010702	Espinho
0109	Santa Maria da Feira	010935	União das freguesias de Santa Maria da Feira, Travanca, Sanfins e Espargo
0205	Beja	020520	União das freguesias de Beja (Salvador e Santa Maria da Feira)
0205	Beja	020521	União das freguesias de Beja (Santiago Maior e São João Baptista)
0206	Castro Verde	020606	União das freguesias de Castro Verde e Casével
0312	Vila Nova de Famalicão	031250	União das freguesias de Antas e Abade de Vermoim
0402	Bragança	040209	Castro de Avelãs
0402	Bragança	040240	Samil
0402	Bragança	040257	União das freguesias de Sé, Santa Maria e Meixedo
0502	Castelo Branco	050205	Castelo Branco
0603	Coimbra	060312	Cernache
0603	Coimbra	060318	Santo António dos Olivais
0603	Coimbra	060333	União das freguesias de Assafarge e Antanhol
0603	Coimbra	060334	União das freguesias de Coimbra (Sé Nova, Santa Cruz, Almedina e São Bartolomeu)
0603	Coimbra	060335	União das freguesias de Eiras e São Paulo de Frades
0603	Coimbra	060336	União das freguesias de Santa Clara e Castelo Viegas
0603	Coimbra	060338	União das freguesias de São Martinho do Bispo e Ribeira de Frades
0604	Condeixa-a-Nova	060411	União das freguesias de Condeixa-a-Velha e Condeixa-a-Nova
0604	Condeixa-a-Nova	060412	União das freguesias de Sebal e Belide
0605	Figueira da Foz	060512	Tavarede
0605	Figueira da Foz	060520	Buarcos
0615	Soure	061506	Granja do Ulmeiro
0705	Évora	070515	Canaviais
0705	Évora	070522	União das freguesias de Bacelo e Senhora da Saúde
0705	Évora	070524	União das freguesias de Malagueira e Horta das Figueiras
0805	Faro	080506	Montenegro
0805	Faro	080508	União das freguesias de Faro (Sé e São Pedro)
0907	Guarda	090758	Guarda
1006	Caldas da Rainha	100617	União das freguesias de Caldas da Rainha - Nossa Senhora do Pópulo, Coto e São Gregório
1009	Leiria	100933	União das freguesias de Leiria, Pousos, Barreira e Cortes
1009	Leiria	100937	União das freguesias de Parceiros e Azoia
1101	Alenquer	110119	União das freguesias de Alenquer (Santo Estêvão e Triana)
1102	Arruda dos Vinhos	110202	Arruda dos Vinhos
1105	Cascais	110501	Alcabideche
1105	Cascais	110506	São Domingos de Rana
1105	Cascais	110507	União das freguesias de Carcavelos e Parede
<b>1105</b>	<b>Cascais</b>	<b>110508</b>	<b>União das freguesias de Cascais e Estoril</b>
1106	Lisboa	110601	Ajuda
1106	Lisboa	110602	Alcântara
1106	Lisboa	110608	Benfica
1106	Lisboa	110610	Campolide
<b>1106</b>	<b>Lisboa</b>	<b>110611</b>	<b>Carnide</b>
<b>1106</b>	<b>Lisboa</b>	<b>110618</b>	<b>Lumiar</b>
1106	Lisboa	110633	Olivais
<b>1106</b>	<b>Lisboa</b>	<b>110639</b>	<b>São Domingos de Benfica</b>
<b>1106</b>	<b>Lisboa</b>	<b>110654</b>	<b>Alvalade</b>
<b>1106</b>	<b>Lisboa</b>	<b>110655</b>	<b>Areeiro</b>
1106	Lisboa	110656	Arroios
<b>1106</b>	<b>Lisboa</b>	<b>110657</b>	<b>Avenidas Novas</b>
<b>1106</b>	<b>Lisboa</b>	<b>110658</b>	<b>Belém</b>
<b>1106</b>	<b>Lisboa</b>	<b>110659</b>	<b>Campo de Ourique</b>
<b>1106</b>	<b>Lisboa</b>	<b>110660</b>	<b>Estrela</b>
<b>1106</b>	<b>Lisboa</b>	<b>110662</b>	<b>Parque das Nações</b>
1106	Lisboa	110663	Penha de França
1106	Lisboa	110664	Santa Clara



<b>1106</b>	<b>Lisboa</b>	<b>110666</b>	<b>Santo António</b>
1107	Loures	110707	Loures
1107	Loures	110726	União das freguesias de Moscavide e Portela
1107	Loures	110727	União das freguesias de Sacavém e Prior Velho
1107	Loures	110730	União das freguesias de Santo António dos Cavaleiros e Frielas
1109	Mafra	110902	Carvoeira
1109	Mafra	110906	Ericeira
1109	Mafra	110909	Mafra
1109	Mafra	110911	Milharado
1109	Mafra	110921	União das freguesias de Malveira e São Miguel de Alcainça
1109	Mafra	110922	União das freguesias de Venda do Pinheiro e Santo Estêvão das Galés
1110	Oeiras	111002	Barcarena
1110	Oeiras	111012	União das freguesias de Algés, Linda-a-Velha e Cruz Quebrada-Dafundo
1110	Oeiras	111013	União das freguesias de Carnaxide e Queijas
1110	Oeiras	111014	União das freguesias de Oeiras e São Julião da Barra, Paço de Arcos e Caxias
1111	Sintra	111125	União das freguesias de Massamá e Monte Abraão
1111	Sintra	111128	União das freguesias de Sintra (Santa Maria e São Miguel, São Martinho e São Pedro de Penaferrim)
1114	Vila Franca de Xira	111413	União das freguesias de Alverca do Ribatejo e Sobralinho
1114	Vila Franca de Xira	111415	União das freguesias de Póvoa de Santa Iria e Forte da Casa
1115	Amadora	111512	Alfragide
1116	Odivelas	111603	Odivelas
1116	Odivelas	111610	União das freguesias de Ramada e Caneças
1214	Portalegre	121411	União das freguesias da Sé e São Lourenço
1306	Maia	130601	Águas Santas
1306	Maia	130608	Milheirós
1306	Maia	130609	Moreira
1306	Maia	130613	São Pedro Fins
1306	Maia	130616	Vila Nova da Telha
1306	Maia	130617	Pedrouços
1306	Maia	130618	Castêlo da Maia
1306	Maia	130619	Cidade da Maia
1306	Maia	130620	Nogueira e Silva Escura
1308	Matosinhos	130812	União das freguesias de Matosinhos e Leça da Palmeira
1308	Matosinhos	130814	União das freguesias de São Mamede de Infesta e Senhora da Hora
1312	Porto	131202	Bonfim
1312	Porto	131210	Paranhos
1312	Porto	131211	Ramalde
<b>1312</b>	<b>Porto</b>	<b>131216</b>	<b>União das freguesias de Aldoar, Foz do Douro e Nevogilde</b>
1312	Porto	131218	União das freguesias de Lordelo do Ouro e Massarelos
1317	Vila Nova de Gaia	131701	Arcozelo
1317	Vila Nova de Gaia	131709	Madalena
1317	Vila Nova de Gaia	131717	São Félix da Marinha
1317	Vila Nova de Gaia	131726	União das freguesias de Gulpilhares e Valadares
1317	Vila Nova de Gaia	131727	União das freguesias de Mafamude e Vilar do Paraíso
1401	Abrantes	140120	União das freguesias de Abrantes (São Vicente e São João) e Alferrarede
1405	Benavente	140503	Santo Estêvão
1410	Entroncamento	141002	Nossa Senhora de Fátima
1416	Santarém	141633	União das freguesias de Santarém (Marvila), Santa Iria da Ribeira de Santarém, Santarém (São Salvador) e Santarém (São Nicolau)
1419	Torres Novas	141920	União das freguesias de Torres Novas (Santa Maria, Salvador e Santiago)
1420	Vila Nova da Barquinha	142001	Atalaia
1420	Vila Nova da Barquinha	142006	Vila Nova da Barquinha
1502	Alcochete	150201	Alcochete
1502	Alcochete	150203	São Francisco
1503	Almada	150303	Costa da Caparica
1503	Almada	150312	União das freguesias de Almada, Cova da Piedade, Pragal e Cacilhas
1503	Almada	150314	União das freguesias de Charneca de Caparica e Sobreda
1504	Barreiro	150407	Santo António da Charneca
1504	Barreiro	150410	União das freguesias de Barreiro e Lavradio
1504	Barreiro	150411	União das freguesias de Palhais e Coina
1506	Moita	150608	União das freguesias de Gaio-Rosário e Sarilhos Pequenos
1507	Montijo	150710	União das freguesias de Montijo e Afonsoeiro
1508	Palmela	150802	Palmela
1508	Palmela	150804	Quinta do Anjo
1509	Santiago do Cacém	150907	Santo André
1509	Santiago do Cacém	150912	União das freguesias de Santiago do Cacém, Santa Cruz e São Bartolomeu da Serra
1510	Seixal	151005	Corroios
1510	Seixal	151006	Fernão Ferro

1511	Sesimbra	151101	Sesimbra (Castelo)
1511	Sesimbra	151103	Quinta do Conde
1512	Setúbal	151207	Gâmbia-Pontes-Alto da Guerra
1512	Setúbal	151209	União das freguesias de Azeitão (São Lourenço e São Simão)
1512	Setúbal	151210	União das freguesias de Setúbal (São Julião, Nossa Senhora da Anunciada e Santa Maria da Graça)
1513	Sines	151301	Sines
1609	Viana do Castelo	160905	Areosa
1609	Viana do Castelo	160948	União das freguesias de Viana do Castelo (Santa Maria Maior e Monserrate) e Meadela
1714	Vila Real	171404	Arroios
1714	Vila Real	171409	Folhadela
1714	Vila Real	171414	Lordelo
1714	Vila Real	171415	Mateus
1714	Vila Real	171420	Parada de Cunhos
1714	Vila Real	171432	União das freguesias de Borbela e Lamas de Olo
1714	Vila Real	171433	União das freguesias de Constantim e Vale de Nogueiras
1714	Vila Real	171438	União das freguesias de Vila Real (Nossa Senhora da Conceição, São Pedro e São Dinis)
1805	Lamego	180525	Lamego (Almacave e Sé)
1823	Viseu	182301	Abraveses
1823	Viseu	182318	Mundão
1823	Viseu	182321	Ranhados
1823	Viseu	182323	Rio de Loba
1823	Viseu	182339	União das freguesias de Repeses e São Salvador
1823	Viseu	182341	União das freguesias de Viseu
3103	Funchal	310303	Funchal (Santa Luzia)
3103	Funchal	310307	São Martinho
3103	Funchal	310308	Funchal (São Pedro)
3103	Funchal	310310	Funchal (Sé)
3108	Santa Cruz	310803	Canico
4203	Ponta Delgada	420312	Ponta Delgada (São Sebastião)
4203	Ponta Delgada	420313	Ponta Delgada (São José)
4203	Ponta Delgada	420317	Rosto do Cão (Livramento)
4203	Ponta Delgada	420325	Santa Clara
4205	Ribeira Grande	420506	Pico da Pedra
4301	Angra do Heroísmo	430103	Angra (Santa Luzia)
4301	Angra do Heroísmo	430116	São Mateus da Calheta
4302	Vila da Praia da Vitória	430203	Cabo da Praia
4302	Vila da Praia da Vitória	430204	Fonte do Bastardo
4302	Vila da Praia da Vitória	430211	Porto Martins
4701	Horta	470108	Horta (Matriz)

Fonte: Código da divisão administrativa (distritos/municípios/freguesias), 2017

É de notar que as freguesias dos municípios de Bragança, Maia, Vila Real e Vila da Praia da Vitória correspondem a quatro regiões, respeitando o município a que pertencem. Além disso, a negrito encontram-se destacadas as treze freguesias comuns aos quatro estudos.

### Anexo C – Apresentação de resultados por freguesia

Figura 30 - Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 1<sup>24</sup>

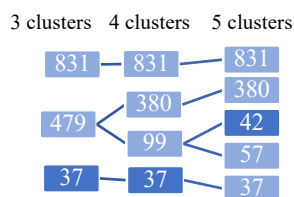


Figura 31 - Percentil 80 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 1, por freguesia, 2016

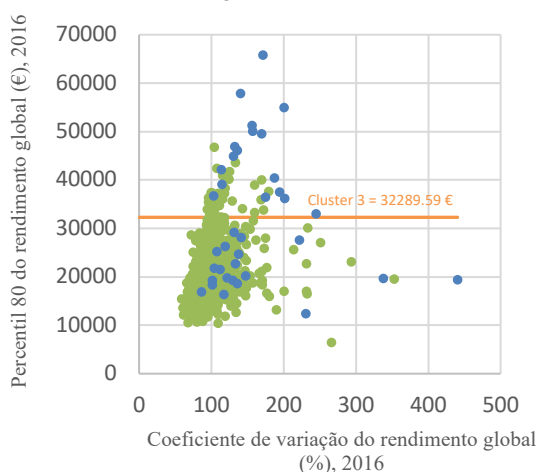
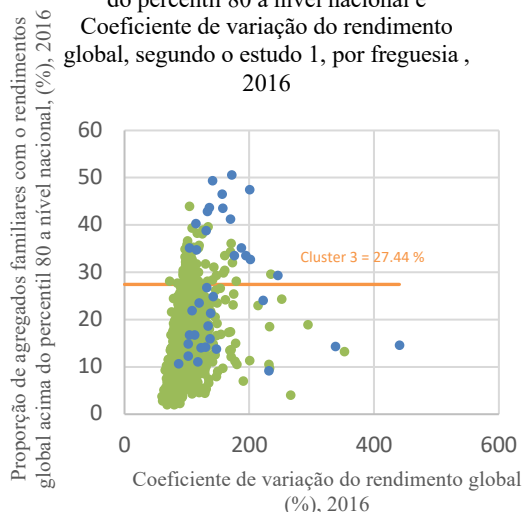


Figura 32 - Proporção de agregados familiares com o rendimento global acima do percentil 80 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 1, por freguesia, 2016



Fonte: Informação associada às notas de liquidação do IRS, 2016

Figura 33 - Percentil 80 dos preços por m<sup>2</sup> de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por freguesia, 2017

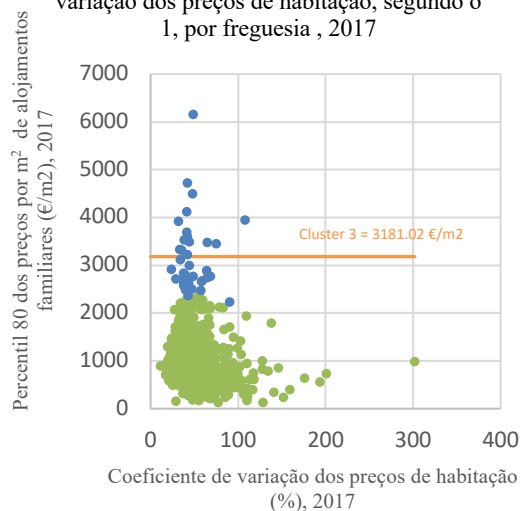
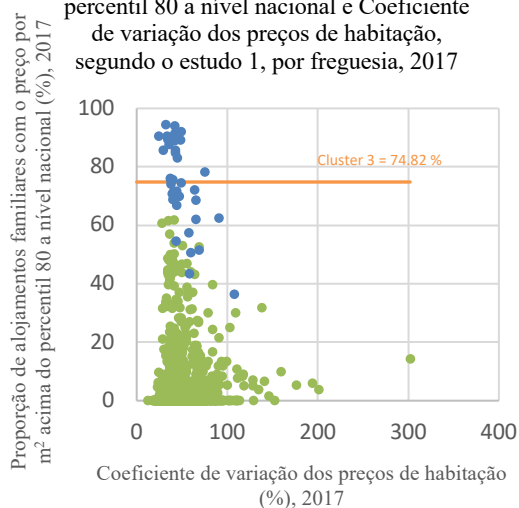


Figura 34 - Proporção de alojamentos familiares com o preço por m<sup>2</sup> acima do percentil 80 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 1, por freguesia, 2017



Fonte: Estatísticas dos preços da habitação, 2017

<sup>24</sup> Foi destacado (a azul escuro) o grupo com maior riqueza com base no rendimento.

Figura 35 - Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 2

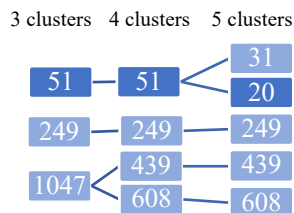


Figura 36 - Percentil 90 do rendimento global e Coeficiente de variação do rendimento global, segundo o estudo 2, por freguesia, 2016

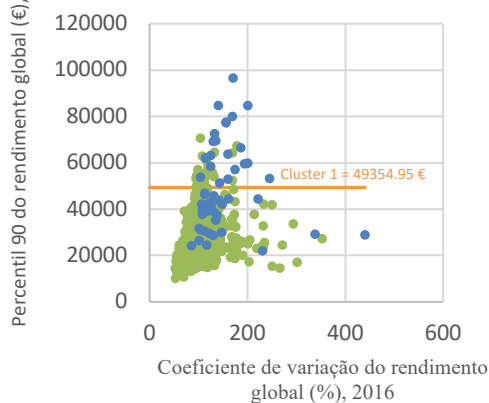
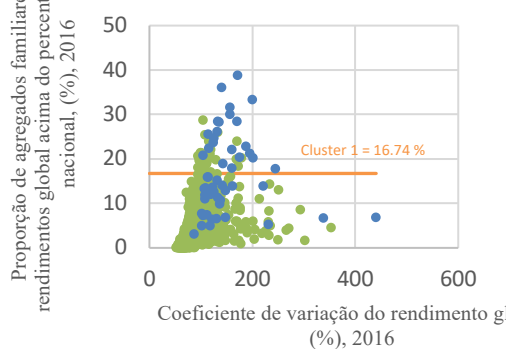


Figura 37 - Proporção de agregados familiares com o rendimento global acima do percentil 90 a nível nacional e Coeficiente de variação do rendimento global, segundo o estudo 2, por freguesia, 2016



Fonte: Informação associada às notas de liquidação do IRS, 2016

Figura 38 - Percentil 90 dos preços por m² de alojamentos familiares e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por freguesia, 2017

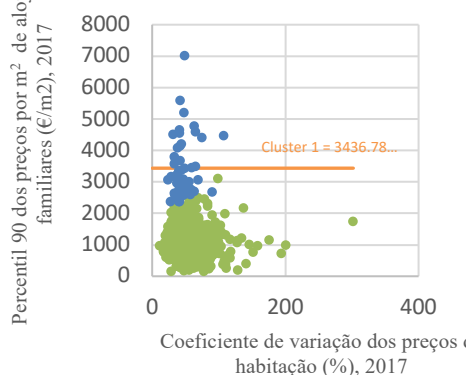
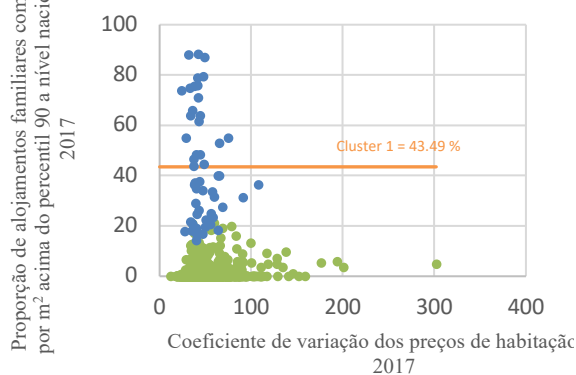
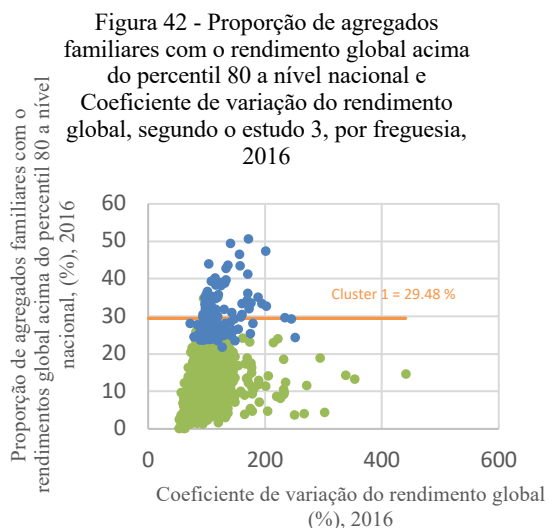
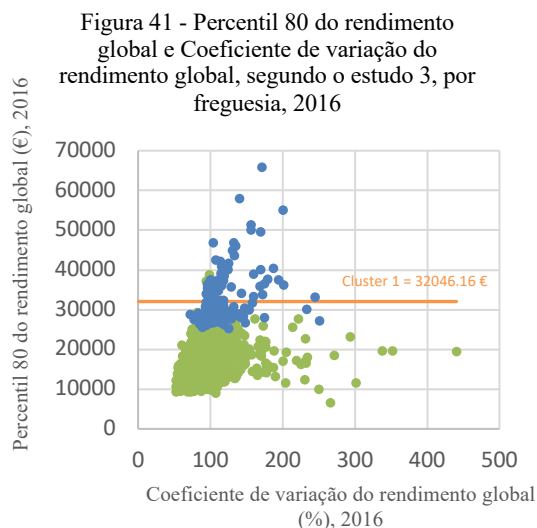
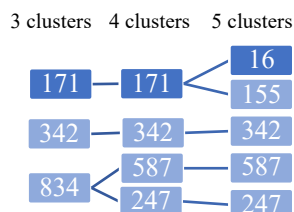


Figura 39 - Proporção de alojamentos familiares com o preço por m² acima do percentil 90 a nível nacional e Coeficiente de variação dos preços de habitação, segundo o estudo 2, por freguesia, 2017

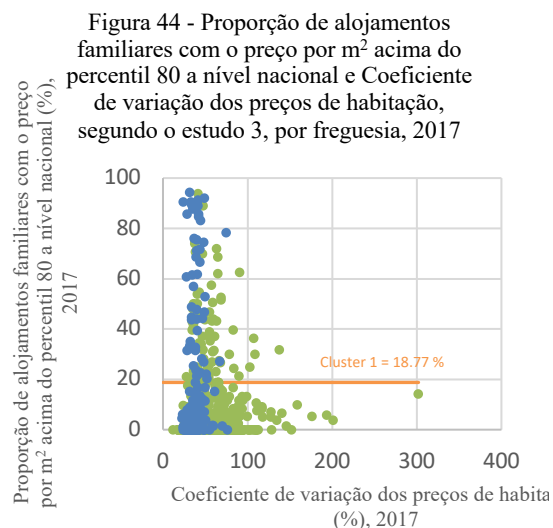
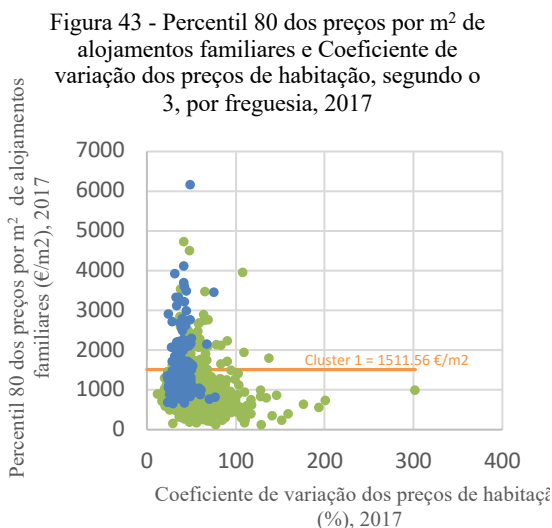


Fonte: Estatísticas dos preços da habitação, 2017

Figura 40 - Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 3

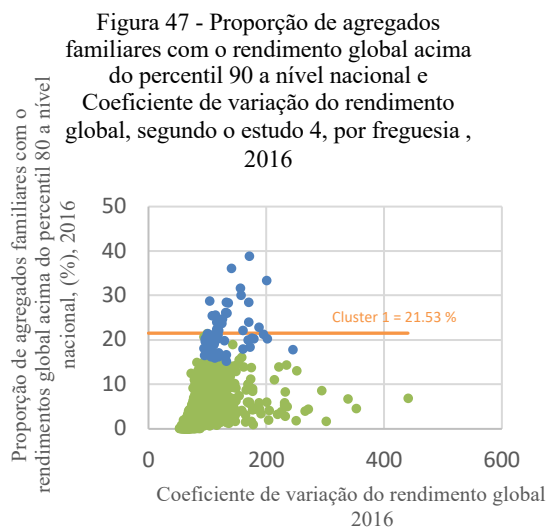
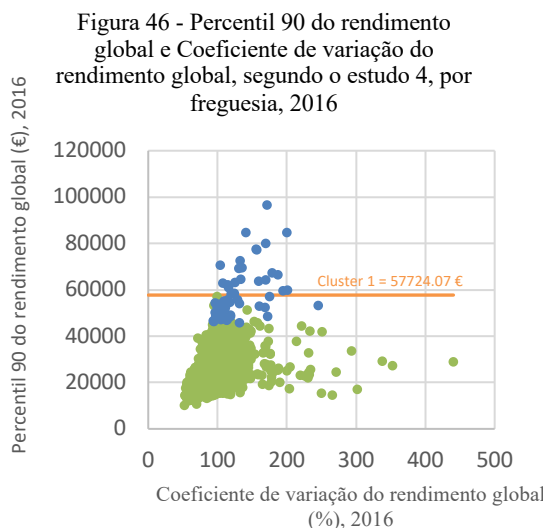
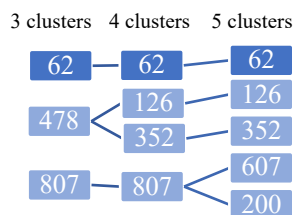


Fonte: Informação associada às notas de liquidação do IRS, 2016

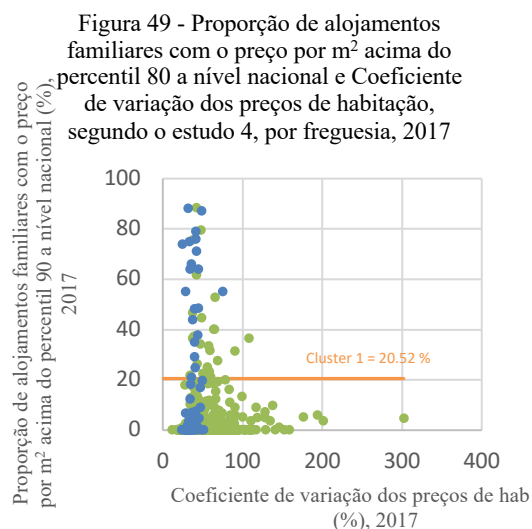
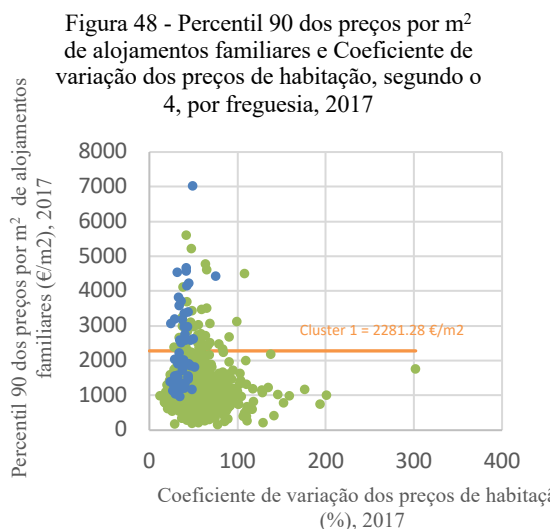


Fonte: Estatísticas dos preços da habitação, 2017

Figura 45 - Agrupamento dos elementos para 3, 4 e 5 clusters do estudo 4



Fonte: Informação associada às notas de liquidação do IRS, 2016



Fonte: Estatísticas dos preços da habitação, 2017

Figura 50 – Classificação das freguesias para três clusters segundo o Estudo 1

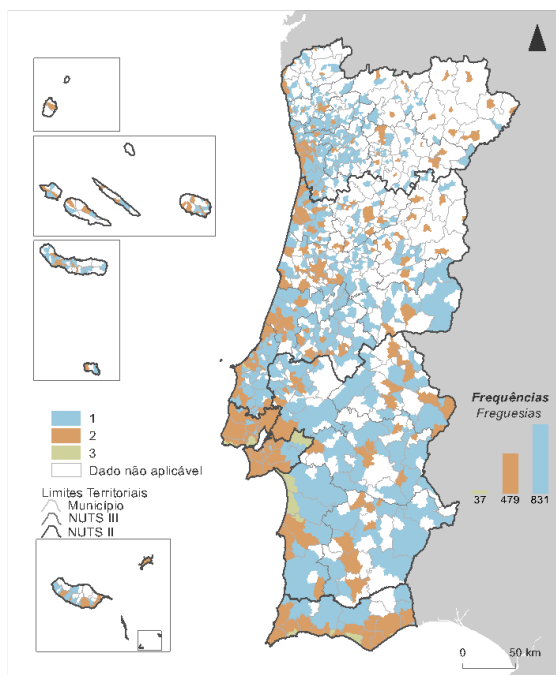
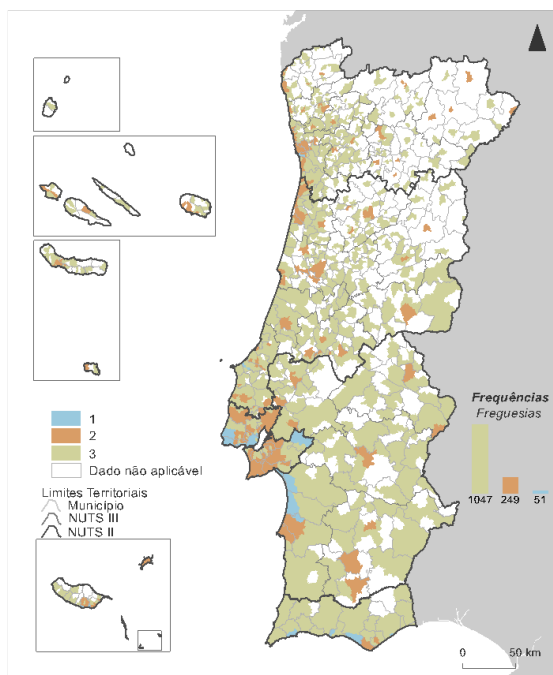


Figura 51 – Classificação das freguesias para três clusters segundo o Estudo 2



Fonte: Informação associada à nota de liquidação do IRS e Estatísticas dos preços de habitação, 2016 e 2017

Figura 52 – Classificação das freguesias para três clusters segundo o Estudo 3

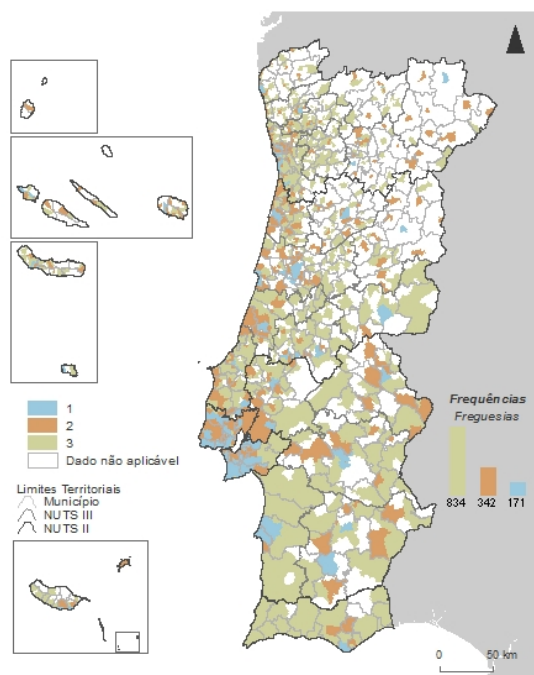
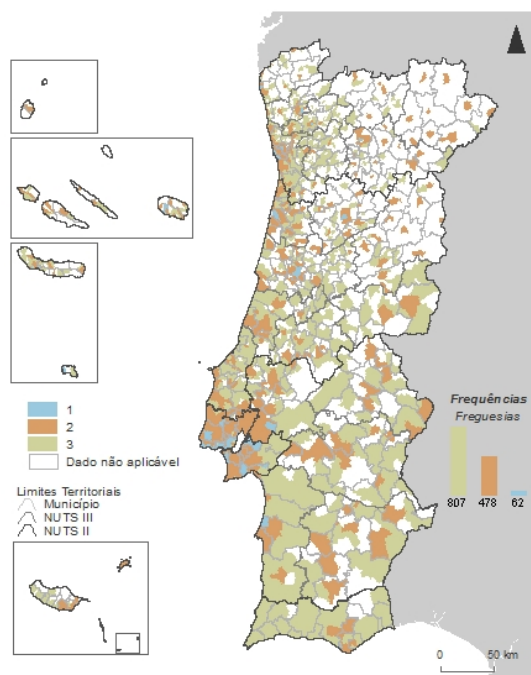


Figura 53 – Classificação das freguesias para três clusters segundo o Estudo 4



Fonte: Informação associada à nota de liquidação do IRS e Estatísticas dos preços de habitação, 2016 e 2017